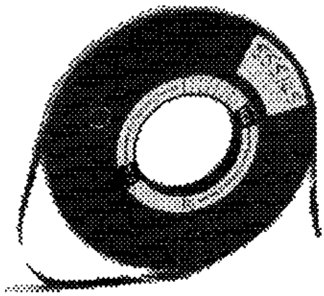


Soil Information Systems Technology

Proceedings of the Sixth Meeting of the ISSS Working Group on Soil Information Systems Bolkesjø, Norway 28 February - 4 March 1983

P.A. Burrough and S.W. Bie (Eds)

Pudoc Wageningen 1984

CIP-gegevens

Soil

Soil information systems technology : proceedings of the sixth meeting of the ISSS Working Group on Soil Information Systems, Bolkesjø, Norway, 28 February - 4 March 1983 / P.A. Burrough and S.W. Bie (eds). - Wageningen : Pudoc
ISBN 90-220-0854-1
SISO 631.2 UDC 681.3:631.4
Trefw.: bodemkunde ; informatiesystemen.

ISBN 90-220-0854-1

Printed in the Netherlands

Contents

Foreword

The sixth fully independent meeting of the Working Group on Soil Information Systems of the International Soil Science Society was held in Bolkesjø, Norway from 28 February to 4 March 1983. Others have been held in Wageningen (1975), Sofia (1977), Canberra (1976, 1980) and Paris (1981). This meeting was organised jointly by the Norwegian Society of Soil Science and the Norwegian Computing Center. Thanks to financial assistance from UNEP, the meeting attracted participants from Bulgaria, Denmark, Federal Republic of Germany, France, India, Kenya, Mexico, the Netherlands, New Zealand, Norway, Poland, Sweden, Turkey, United Kingdom, USA, USSR, Venezuela, Yugoslavia, and Zambia, making it the most international meeting of the Working Group to date.

As the contents of this volume reveal, the interests of the Working Group have broadened since its inception in 1974 to include not only soil database management, but also applications in geostatistics and spatial analysis, in land evaluation and in the new possibilities opened up by microcomputers and remote sensing. It is to be expected that activity and interest in all these fields will continue to develop as soil information systems become more easily accessible. As people begin to realise that 'computerizing' can mean more than just going over from paper to electronic data storage and retrieval, but can also provide powerful tools for analysis and synthesis, we may expect many new insights and developments in the fundamental understanding of soil science and its applications.

the Editors

Soil data in digital space[1]

Stein W. Bie, Norwegian Computing Center, P.O. Box 335 Blindern, Oslo 3, Norway

Dimensions in soil science

Among the many sciences, soil science has reached maturity. The challenge is now to prevent our science becoming overripe. In this paper, I wish to suggest that one key to further development in soil science lies in a revision of the set of dimensions on which it is now constructed. The limit of our perception of space is three. But we know well that in a mathematical sense space can take on any dimension. Soil science has had an uneasy relation with mathematics, and so far relied on simple two- and three dimensional tools for information transfer. They have served us well. It is not with disregard for past achievements that we may now investigate new dimensions.

There are three 'dimensions' of soil science that I believe to be key areas to progress. In this paper, I wish to elaborate on each of the three 'dimensions' and offer my view on the road ahead. Dimension one is position: the accuracy of positioning of observations and data. Dimension two is covariance: the relation between simple soil data and the real-world tasks for which we collect soil data. Dimension three is information transfer: the vehicle we employ to transport our soil science world to the community outside.

In conclusion, I shall explore the bridge between the low dimensionality of our present soil science and the high dimensionality that I think will be associated with the new.

Dimension one: position

Many geological and soil survey organizations combined the job of topographic survey with their more direct earth science tasks. The need for positional referencing of soil data created a need for a geodetic framework.

In many countries of the world, the lack of geodetic referencing is a major handicap for the progress of soil science. Soil can change over very short distances. If we are unable to define a

1. An extended version of this paper was presented in Dutch at the Seminar in honour of Dr J. Schelling on 21 April 1983 as 'Bodemgegevens in digitale ruimte'. The Dutch version is published in: Bodem en landschap kwalitatief en kwantitatief bekeken, E.C.W.M. Ruyter et al. (Eds). Pudoc, Wageningen, 1984. p. 36-44.

point of reference, we cannot relate the soil data to the world around.

The geodetic problem is severe in the developing world. We know that the official geodetic referencing system of Africa has geodetic referencing points that are probably as much as 10 km wrong. That is 10 000 m, in a soil world where 50-80 % of all soil variability may take place within a few hundred metres. In a paradoxical way, the Landsat satellites intended for soil use (and disappointing so many soil scientists) have rewarded many of the Third World countries most richly in their geodetic accuracy, for which they were not intended. Even based only on orbital parameters, the geodetic accuracy of Landsat images exceeds conventional maps by a factor of 5 to 100 in the remoter parts of our world. Although unrelated to soil, the navigational satellites of the Transit generation and now Navstar can focus the x,y coordinates of soil observations. It is a major challenge to the space industry to bring down the cost of satellite positioning equipment from US$150 000 now to, say, US$3000 per receiving set, and thus give us the chance to pinpoint a soil observation anywhere in the world to 5 m. In the long run, we should not only offer soil maps to the developing countries but soil maps with a sound geodetic basis. If - like diamonds - soil maps are for ever - they had better get the position right.

A tragedy for soil science is the military straitjacket put on many of our colleagues in reporting positional detail of soil. It is incomprehensible that superpowers able to drop bombs on each other with metre accuracy are unwilling to allow the publication of soil maps with greater geodetic accuracy than old bathroom wallpaper. The inability to relate soil to topography by maps is likely to deprive our colleagues in many countries of the simplest yet essential dimension in their work.

For many parts of the world, an ability to pinpoint soil data or the permission to do so, is - I believe - the greatest single contribution to soil science.

So dimension one is really two-dimensional: grant us accurate x and y coordinates.

Dimension two: covariance

There are some soil data that have direct relevance to use: soil engineering data constitute a (nearly) complete data set for the construction of roads. Most collectors of soil data are less fortunate. For those of us with roots in agriculture and forestry we are well aware of two things:
- soil data constitute only one of many factors responsible for productivity and suitability, and not necessarily the dominant one;
- we often collect data about soil variables that are easy or cheap to measure, in the hope that these variables are strongly correlated with those that are more directly related to productivity,

or are very expensive to measure.

There are those of us who have repeatedly pointed to data on farm and forest economics showing that soil factors - in the manner in which we have measured them - have a varied and often weak influence on the variability of production. A study on world food production is about to be published by the International Institute for Systems Analysis in Vienna, on contract to FAO. As soil scientists, we will be sad to discover the minor role they see soil play in the world food production scenario, and the great emphasis given to climate, management skills and economic infrastructure. As soil scientists, we have two ways of handling the seemingly diminishing role of soil information:

- as we have done in the past: ignoring the economists and just getting on with our soil science;
- taking up the challenge, and encouraging independent sound research on soil and productivity relations.

The work on land evaluation and land/soil productivity seems to me to have entered a cul-de-sac. Its strong physiographic and soil bias and the fact that so many of its adherents stick to the physical features of the landscape are endangering the role land evaluation plays both in the developed and the developing world. Complex land evaluation models, full of subjective weightings and value transformations, are no longer credible. Many here in the audience will by now be thoroughly on the defensive, but they, too, should realize that every approach has its time and place. A tool may seem optimum at a particular time, but methods and economics roll on. Practising agriculture and forestry in the shadow of the present oil price is very different from that of ten years ago. It is a tenable hypothesis that the recent spread of deserts is as much a function of the high oil price and the ensuing struggle for cheap wood and charcoal energy, as of any climatic or demographic factors that we may suggest. At least, I find it hard to think of soil factors.

Disappointingly, many seem also to reject the once promising mathematical models of crop production. The bias towards biology seems to cut little ice among planners and the financing community. If soil science wishes to come into closer contact with executing agencies involved in planning, both in the developing world and in industrial countries, some of us must devote time and intellectual effort towards assessing the role of soil data in planning. Working with international planning agencies over the last few years, I have been struck by the timidity of soil scientists to demonstrate the value of soil data.

The period of environmental euphoria is over, but environmental concern persists! At last Europe and North America are waking up to acid rainfall, which has been associated with the dying lakes of Scandinavia for the past 25 years. Last summer, the first Scottish lakes turned sour - all 27 of them having conifers planted in their watersheds about ten years ago. Suddenly last summer

German spruce trees dropped their needles. We are facing complex environmental destruction, where atmospheric gases, both sulphurous and nitrogenous, interact with vegetation and land-use variables and soil cations and biological aspects. When acid rainfall first became an issue in Scandinavia, we overlooked the complexity of the soil. We paid a heavy price: political disbelief among those countries that emit most industrial pollution. We can count the price today in Norway in hundreds of dead lakes. And the count has just begun in Britain, in Germany, in Canada and in the United States. Soil is not everything, but it is one dimension. Soil scientists must be prepared to work actively in an interdisciplinary setting, and traditional soil organizations must not be afraid of other disciplines.

The other type of covariance is statistically easier. Some soil variables are easy or cheap to measure: Munsell colour, field texture, depths of horizons and layers, rooting depth, mottles and concretions or hydrochloric acid test. So much of our field soil science is tied to such simple measurements. Yet there are hardly any studies published to relate these variables to others that we know are relevant to many kinds of land-use. We replace these covariance studies with reference to classes in soil classifications, expecting soil classes to be homogeneous in respect to the other variables. An imaginative engineering firm should build and market a miniature soil-testing laboratory. Looking at the progress made in miniature medical testing facilities, I judge it within the reach of modern know-how to construct an automated soil laboratory the size of a small suitcase. And I am sure there would be a market for it. Let it be said that I interpret the lack of major studies on covariance between soil properties on local, regional, national or global scale to suggest that there is very little of such covariance.

The covariance argument is central to any discussion on derived or interpreted soil maps. In many countries the translation of a general soil map into a derived map has been a recolouring of the soil map, with no new boundaries added: only some old ones deleted. On a global scale, we have had a particularly blatant case recently, when the UNESCO/FAO Soil Map of the World was recoloured to give actual soil degradation and risk of soil degradation in North Africa and the Middle East. Valuable as this interpretation of soil maps may have been in the past and vital as it may seem to the commissioned survey departments of soil survey organizations, the time has come to have a deep rethink on interpretation of soil data for applications. I now wish to suggest a strategy for this rethink.

Interpretation is a set of algorithms that we apply to a model. The quality of the result is a function of both the underlying model and the algorithms. In the present production of interpreted maps, we have erected a train of thought that runs like this:
1. We know some soil properties relevant for a particular land-use.
2. We grade the classes in our soil classification according to their values for these properties.

3. We average the grade for mapping units by calculating the proportion of each soil class in each unit.
4. We recolour the soil map according to the calculated average grade.

Step 1 to 2 relies on absolute faith in strong covariance. Step 2 to 3 relies on the limits of mapping units being commensurate with the limits of soil classes and that the proportion of each type of soil in each unit is known with high accuracy. Step 3 to 4 is mechanical but is based on soil data only.

I wish to argue that this line of thought, useful as it may have been in the past, has reached its limits of resolution and that a new and stronger magnification must be put into our 'soil microscope'.

I suggest that Steps 2, 3 and 4 be dropped, to leave us with raw soil data only. This cuts out the need for beliefs in covariance in classification. New data must be added: some are soil variables that are normally not measured and some are non-soil variables: the latter will vary locally but include farming skills, available machinery, economic factors (including the oil price!), policy and management.

I am therefore suggesting that for the interpretation of soil maps both the model and the algorithms be changed. The model should no longer be 'the soil map' but a larger set of soil data plus biological, technical and socio-economic data. The algorithms should no longer consider only soil data but all data. They should not be national but local. The interpretation of soil maps becomes land evaluation, not in a global sense, but optimized locally.

In many countries, the change in interpretative model and in algorithms will cause institutional problems. A soil survey institute was created in an older purer soil world. It was the role of the consulting engineering firms to use the soil information in the final planning process.

With the new model and new algorithms, a soil survey institute will inevitably have ambitions to include areas that the consultancy firms have regarded as theirs. On the other hand, the consultancy firms depend on the soil expertise for their models. We are seeing clearly how scientific changes may have organizational and institutional consequences. There are many ways of solving these. One thing is at least universal: in the long run, it is silly to erect artificial breaks in the information handling chain in order to serve a dated institutional set-up. Sweden has recently had this problem in remote sensing, and chosen a 100 % state owned commercial corporation (Swedish Soace Corporation) to handle all aspects of remote sensing, from satellite launches to interpreted maps. However, private firms are free to choose niches in the total field and may enter into normal commercial deals with the state firm. It is a model worth looking at also in the context of soils.

Thus the study of covariance has taken us to the institutional consequences of integrating soil data with other data. Time has come to consider the third dimension.

Dimension three: information transfer

The preoccupation in the past with data technology has sometimes obscured the search for inherent data transfer functions, i.e. tools for retrieving and enriching information. Now that data technology is becoming more common-place, we have time to reflect.

Fundamental to our understanding of soil is information about soil properties. Soil is classified according to soil properties. Soil map legends are constructed on the basis of soil classes. Map legends yield mapping units. Soil surveyors delineate occurrences of these mapping units. However, soil scientists differ when they collect sets of soil properties for a point. There are differences in emphasis on alternative properties, differences in the definitions of layers and horizons, and in the field and analytical methods used to acertain their values. In addition some surveyors are better at estimating values than others, and some laboratories are better equipped or more careful than others.

Soil surveyors differ significantly in their choice of soil classification. In most countries, several classifications have been used over the years, and to varying depths of subdivision. The ability of individual surveyors to use the classifications have varied too. I am therefore suggesting that soil archives of point data contain sets of data of inherently heterogeneous nature. A further complicating matter is the spatial sampling procedures used, varying from obtaining 'the typical' or 'the representative' to a statistical random or stratified sample. Map legends are partly tied to soil classifications, but legend units may contain one or more classes. In practice, mapping units may contain one or more legend units (e.g. complexes or associations). In the delineation of occurrences of mapping units in the field, on air photographs or other remotely sensed images, or by statistical means (e.g. interpolation), further inaccuracies are inevitable.

A compounding factor is map scale, strongly influencing the choice of legend, mapping units and the ability to convey complex or intricate occurrences. A final complication is caused by dimension time, where soil properties change with time (seasons, cultivation, practices, natural disasters, pedological soil development). Although frequently overlooked by soil scientists (particularly those of Western European tradition), time changes soil.

In the world, soil data is a heterogeneous collection, seemingly impossible to standardize retrospectively, yet a tremendous inheritance if we could obtain a tool of insight not restricted by changing traditions and practices.

Recently such a tool has been proposed (Bie & Lamp, 1983), although in a slightly different context. It is only fair to stress that it is an untried approach, but it contains interesting ingredients suited to tackling the soil problem. The approach, involves projecting soil classes in existing soil classifications back into an n-dimensional hyperspace where they are parameterized and thus comparable. I will now discuss the methodology in some detail:

Using the definition of a soil class

A soil class (in any traditional soil classification) is described by a set of definitive criteria. Each definitive criterion (soil property) has a minimum and a maximum value for each class. If there is a set of n-definitive criteria, the soil class is described by a set of n minimum and maximum values. This set is called the class envelope and occupies a unique position in the n-dimensional hyperspace. Other classes occupy other unique positions in either n-dimensional hyperspace or in an m-dimensional hyperspace (as all classes need not be defined on the same set of definitive criteria). In a given soil classification no two classes can occupy the same position in a hyperspace defined by the smallest set of criteria used for one class; i.e.: there are no overlapping classes. This property of a soil classification enables us to say specific things about a classified soil individual (e.g. profile). The narrower the range of values for each definitive criterion, the higher the precision of statements we can make. If we know a soil individual by its class name only, we can deduce its possible minimum and maximum value for a definitive criterion by referring to the class envelope.

Improving on a class envelope estimate

If we have a large number of soil individuals classified to belong to one class, we may expect that these individuals lie scattered around within the class envelope. We can measure this scatter and find its centre of gravity within the class: its centroid. It often happens that many soil individuals lie close to the centroid, within a subenvelope of the class envelope. This envelope is described by a set of narrower ranges of values than the class envelope. If p percent of all individuals in one class lie within the subenvelope, there is a p percent probability that a soil individual has these narrower ranges of values. If p is large, we can further improve on the precision of our estimate by using the estimate of the subenvelope.

The non-definitive soil properties

In addition to the criteria used to define the soil class, a soil individual may have a further large set of observed or measured soil properties. The purpose of a general soil classification is to ensure that a soil class is also reasonably homogeneous in respect to these associated soil properties. This is the 'covariance problem'. The utility of a soil classification depends on the homogeneity of a soil class in respect to the soil properties (definitive or associated) important for a particular soil use. This we discussed under Section Dimension two.

We can now include a number of associated soil properties alongside the definitive criteria, and define an extended class envelope in s-dimensional hyperspace, where s is the sum of the definitive and the associated properties. If we include all n definitive criteria the position of the extended class envelope of a class will also occupy a unique position in s-space, and a soil individual may be uniquely classified. However, if we exclude one or more definitive criteria in the definition of an extended class envelope, these envelopes may now be overlapping. This means that a soil individual belonging to one class may find itself within the extended envelope of one, two or more classes.

If we have a large number of classified soil individuals, each with a set of associated properties, we can use these individuals to describe the extended class envelope for that set. As with the definitive criteria we can evaluate whether an extended class subenvelope may describe a large proportion of soil individual within the larger extended envelope.

It is worth noting that extended class envelopes can exist for any set of soil properties.

Alternative soil classifications

The arguments above apply to any traditional soil classification. We can thus project any class of such classification into a multidimensional space and find its class envelope, its subenvelope, its extended envelope and extended subenvelope. Describing a soil individual by its position in multidimensional space enables the comparison of two individuals in that multidimensional space, regardless of the soil classifications with which they have been classified.

Soil classifications differ, however, in their sets of definitive criteria. In order to project two soil individuals classified in different soil classifications into the same multivariate space, we need to supplement either with a number of associated soil properties. There seem to be two ways of doing this:
1. By classifying a reference set of soil individuals into classes in a number of relevent soil classifications and measuring a large set of associated properties of each, to create a definitive reference base.
2. By matching soil individuals that are similar in a smaller number of properties to estimate their values for other attributes for which some of them have values and others not, to create an estimated reference base.
A reference base may consist of mixtures of definitive and estimated data.

Applicability to soil map legends and soil mapping units

A soil map legend defines the kinds of soil present in an area in terms of mapping units. The legends may be of differing kinds, representing various paths to information at increasing levels. Map legends are defined in terms of mapping units, which again are defined in terms of classes in soil classifications. For simple map legends with mapping units containing one soil class only, we can make an estimate of the values of soil properties in the delineated area by referring to the national reference base of that class. This estimate is subject to many errors (among these map impurities), but may be the best available. If supplemented by local soil information frequently provided in the survey memoir, the estimate may be further improved.

Reference bases

The preceding paragraphs can be used to argue for the establishment of national and international reference bases. In the past it has been inconceivable to establish such bases due to the very large number of data such bases will inevitably contain. It is possible to establish reference bases by sampling all available data, thus reducing the actual effort involved. The actual sampling design used is of great importance to the result, and must be the subject of a separate study evaluating the available material. A reference base is most easily established by using numerical classification techniques and basing the referencing completely on computer-aided methods.

Operating in multi-dimensional space is beyond the task of 'normal' people. Our reliance on computers will therefore increase. Much development in information technology has taken place. The first generation of minicomputers came, now we have the second. But parallel to this is the microrevolution, and possibly more important to us, the design of specialized computer architecture for graphical data handling.

We are almost to a stage where many soil survey organizations may find it worthwhile to invest US$3000 in each staff member by giving him or her his or her own soil micro. Even if half of them stood idle, the return on investment might be considerable. Many soil organizations now have a general change of staff. It is just imperative that the new people are open to new tools of data handling. 'Why', an Israeli friend of mine asked me the other day, 'why did Moses spend 40 years in the desert with his people? After all, it was a relatively short walk across, a few weeks at the most and certainly not 40 years. Moses spent 40 years in the desert because he needed a new generation of his people, a generation that was liberated from past constraints, had no memories of strict rules'. Retaining people is more than filling them with new facts and new skill. The key word is 'new will'. And unlike the former it cannot be bought by money. Only time.

Conclusions

The bridge between the low dimensionality of old soil science through its classifications, interpretation tables and general purpose soil maps and the high dimensionality of the new soil science is paved with statistics, real-life models and electronic data processing. But these paving stones are just the foundation for a science to proceed. We are at last liberated from some of the old constraints that we had learned to live with and liberated from first generation data technology that we love to hate. So soil science can proceed, in digital space, but very much on the ground.

Reference

Bie, S.W. & J. Lamp, 1983. Criteria/soft-hardware for Global Land/ Soil Monitoring System. Version 2. UNEP, Nairobi. Na. 83-5175.

Relational Pascal model of soil database

S. Alagić (IRCA-Energoinvest), M. Ćirić (Faculty of Forestry and B. Zaciragić (Faculty of Electrical Engineering), University of Sarajevo, YU.71000 Sarajevo, Yugoslavia

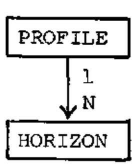
Abstract

The design of a relational model of a soil database is presented. The model contains representations of the profile and the horizon entity types, their properties and the relationships among them. Typical queries upon the soil database are expressed in terms of a particular unified query and programming language which is based on Pascal. The chosen examples demonstrate the advantages of the relational model and the database management system used for its implementation.

Logical database design

Our soil database is a representation of two types of entities and the relationship among them. The first entity type is called profile. Its existence in the database does not depend upon the existence of other entities and because of that the profile is called the kernel entity type. The other entity type is the horizon. The horizon is a characteristic entity type whose existence in the database depends upon the existence of the profile entity to whom it belongs. This classification is performed according to Codd (1979).

The relationship between the profile and the horizon entity types is 1:N. This means that a profile has a finite number of its horizons and that every horizon belongs to one and only one profile. The conventional representation of this relationship is by the following diagram:



The above abstract model is further decomposed in such a way that each of the above entities is represented by three sets of tuples (flat records). (A tuple is another term for a record, meaning a set of attribute descriptions and their values. Eds.) These sets of tuples may be viewed as relations.

The profile entity is represented by the following three sets (relations):

PROFILE	FACTORS	SOIL INDIVIDUUM

The above sets correspond to groups of related attributes. The first group of attributes identifies the profile. The second group of attributes are factors of soil information. The third group contains attributes of the soil individuum.

The horizon entity is represented by the following sets (relations):

HORIZON	MORPHOLOGY	CHEMICAL

The first of the above sets of attributes identifies the horizon, the second contains the attributes which describe the horizon's morphology and the third its chemical properties.

In the next step of the decomposition procedure the attributes in each group are specified by their names. We thus obtain the following model of a soil database.

PROFILE (P, NUMBER, STATE, INSTITUTION, SECTION, SHEET, CO-ORDINATES, TYPE, SUBTYPE, VARIETY)

FACTORS (P, ELEVATION, SLOPE, ASPECT, TOPOGRAPHY, GEOMORPHOLOGY, ROCK, WEATHERING, ACCUMULATION, TEMPERATURE, RAINFALL, STATION, TREES, VEGETATION, AGRICULTURE, WATER, WIND, FLOODS)

SOIL INDIVIDUUM (P, DRAINAGE, PERMEABILITY, STONES, EROSION, SOLUM, CARBONATES, ROOTS)

HORIZON (P, ORD, LOWER, UPPER, KIND, TRANSITION, DISCONTINUITY, DESIGNATION)

MORPHOLOGY (P, ORD, TEXTURE, STRUCTURE, SKELETON, CONSISTENCY, CEMENTATION, POROSITY, COLOUR)

CHEMICAL (P, ORD, REACTION, CARBON, NITROGEN, CARBONATES, PHOSPHORUS, POTASSIUM, EXCHANGE, CATIONS, HYDROGEN)

The attribute P has the role of the primary key which means that this attribute alone identifies the profile entity. The role of the primary key of the horizon entity is performed by two attributes: P, the identification of the profile to which the horizon belongs, and ORD, the ordinal number of the horizon.

The fourth step of the database design would be a specification of the domains of the attributes which appear in the model. The domain of an attribute is a set of values which an attribute may assume. The specification of the domains of the attributes of our database model will not be given in this paper, although specific values will appear in sample queries in the next section. We just mention that the domain of the attribute P is integer and so is the domain of the attribute ORD.

The final design step is the specification of the database in the data definition language. Our database management system has a unified data definition and manipulation language which is based on the programming language Pascal (Schmidt, 1977; Alagić & Kulenović, 1981). As such, it differs considerably from the system used by Kollias & Yassoglov (1981). Rather than getting into various details of the language (which is similar to Pascal-R), we present a series of sample queries which demonstrate its power.

Sample relational queries

The simplest form of a query specifies one set of tuples (relation), a logical condition which selects relevant tuples from that set and the attributes of the selected tuples which are required as the output. So the result of a query is again a set of tuples. An example of such a query is:

Print the ratio of carbon and nitrogen for all horizons with the reaction between 5 and 7 and the percentage of carbon greater than 10.

```
FOREACH X IN CHEMICAL
WHERE (X.REACTION > 5)
AND   (X.REACTION < 7)
AND   (X.CARBON > 10)
AND   (X.NITROGEN <> 0)
DO WRITELN(X.CARBON/X.NITROGEN)
```

The control variable X stands for a tuple of the set CHEMICAL so that X.REACTION, X.CARBON and X.NITROGEN denote values of the attributes REACTION, CARBON and NITROGEN respectively. Only those tuples X from the set CHEMICAL are selected for the output which satisfy the condition (X.REACTION > 5) AND (X.REACTION < 7) AND (X.CARBON > 10) AND (X.NITROGEN <> 0). Finally, the ratio X.CARBON/X.NITROGEN is printed out or displayed at the terminal where the query is typed in.

The next example of a query is more complex since it refers to two sets of tuples.

Print the profile number and the institution for all the profiles in the state of Bosnia with a horizon whose depth is greater than 30 and whose designation is A.

```
FOREACH X IN PROFILE
WHERE  (X.STATE=BOSNIA)
DO  FOREACH Y IN HORIZON
    WHERE (X.P=Y.P)
    AND   ((Y.UPPER-Y.LOWER) > 30)
    AND   (Y.DESIGNATION=A)
    DO  WRITELN(X.NUMBER,X.INSTITUTION)
```

In the above example, one query is nested within another. The outer query is specified upon the set of profiles and the inner upon the set of horizons. The condition which relates horizons to their profile is equality of the values of the key attributes (X.P=Y.P). In the next query, selected horizons are related to their chemical properties.

Print the contents of carbon for all horizons with the designation A, the gradual type of transition and the reaction greater than 6.

```
FOREACH X IN HORIZON
WHERE  (X.DESIGNATION=A)
AND    (X.TRANSITION=GRADUAL)
DO  FOREACH Y IN CHEMICAL
    WHERE (X.P=Y.P)
    AND   (X.ORD=Y.ORD)
    AND   (Y.REACTION > 6)
    DO  WRITELN(Y.CARBON)
```

Observe that the condition which relates horizons to their chemical properties is (X.P=Y.P) AND (X.ORD=Y.ORD).

The next query is by far the most complex of all the examples in this paper. It exhibits the explicit notation for a query specified upon two sets of tuples. Furthermore, such a query is embedded into another query upon two sets of tuples so that we obtain a query upon four sets of tuples. The outer query refers to profiles and their properties and the inner to their horizons and their properties.

Print the value of the reaction of the horizons with the designation A and the exchange less than 30 for all the profiles of the type eutric cambisol whose depth of carbonates is greater than 100.

```
FOREACH X IN PROFILE, Y IN SOIL INDIVIDUUM
WHERE (X.TYPE=EUTRICCAMBISOL)
AND   (X.P=Y.P)
AND   (Y.CARBONATES > 100)
DO  FOREACH Z IN HORIZON, W IN CHEMICAL
    WHERE (Z.P=X.P)
    AND   (Z.P=W.P)
    AND   (Z.ORD=W.ORD)
    AND   (Z.DESIGNATION=A)
    AND   (W.EXCHANGE < 30)
    DO  WRITELN(W.REACTION)
```

One of the major advantages of our system is that it provides a unified query and programming language for its users. Our final example shows how query and procedural facilities are combined in a modest fashion in order to compute a simple statistical value.

Print the mean value of the contents of nitrogen for the first horizon of all profiles whose elevation is between 1000 and 1500 m.

```
BEGIN SUM:=O; N:=O;
      FOREACH X IN FACTORS, Y IN CHEMICAL
      WHERE (X.ELEVATION > 1000)
      AND   (X.ELEVATION < 1500)
      AND   (Y.ORD=1)
      AND   (X.P=Y.P)
      DO  BEGIN SUM:=SUM+Y.NITROGEN; N:=N+1
          END;
      IF N<>O THEN WRITELN('MEAN VALUE OF NITROGEN:';
END.                           SUM/N
```

Implementation remarks

A soil database with 10 000 records has been implemented according to the model described in this paper. The computer was a UNIVAC 1100. This database served as a test database for a newly developed relational database management system based on the programming language Pascal. Perhaps the most important feature of this system is its portability. It is completely written in the programming language Pascal with only 1 % of machine code, so that it can easily be adapted to various computers as long as they have a good Pascal compiler. A number of database queries written in the unified relational query and programming language of this system have been tested with very good response times.

References

Alagić, S & A. Kulenović, 1981. Relational Pascal database interface. Computer Journal 24(2): 112-117.

Cirić, M. & J. Palc, 1977. Development of a soil information system in Yugoslavia. In: Developments in Soil Information Systems. Proceedings of the Second Meeting of the ISSS Working Group on Soil Information Systems, Varna/Sofia, Bulgaria, 30 May-4 June 1977. (Dr Alexander N. Sadowski and Dr Stein W. Bie, Eds), p. 73. Pudoc, Wageningen.

Codd, E.F., 1979. Extending the database relational model to capture more meaning. ACM Transactions on Database Systems 2(3): 247-261.

Cormack, R.S. & A.W. Moore, 1980. Modelling the land resource data for computer storage and retrieval. In: Developments in Soil Information Systems. Proceedings of the Second Australian Meeting of the ISSS Working Group on Soil Information Systems. Canberra, Australia, 19-21 February 1980. (A.W. Moore et al., Eds) pp. 8-24. Pudoc, Wageningen.

Kollias, J.G. & N.J. Yassoglou, 1981. A query language for retrieving information from a soil data bank. Computers and Geosciences 7(4): 393-400.

Schmidt, J.W., 1977. Some high-level language constructs for data of type relation. ACM Transactions on Database Systems 2(3): 247-261.

The ADK data base systems (PDS and LADS) used in Denmark

S.W. Platou, Ministry of Agriculture, Bureau of Land Data,
Enghavevej 2, DK-7100 Vejle, Denmark

Abstract

At the Bureau of Land Data (ADK) an integrated data-processing system is used for handling soil and other natural resource data. The system consists of two main parts, one handling parametric data and one handling line-area data. Since 1978, large data bases have been established within the frames of these systems, used in the regional agricultural planning, water resource planning and land use planning.

Introduction

The database system used at the Bureau of Land Data (ADK), under the Danish Ministry of Agriculture, is a further development of the so-called SYSTEM II (Platou, 1971; 1975). The database handling software is divided into two main systems:
1. The Parameter Data System (PDS, Platou, 1983a) handles all parametric descriptions, such as field observations, sample analyses (e.g. texture, chemical elements), profile descriptions (including a water supply drill-hole data handling subsystem).
2. The Line Area Data System (LADS, Platou, 1983b) handles line-coordinate data. The lines can be boundaries between areas with different properties, e.g. different soil types, or linear elements in nature, e.g. water courses. The system also handles topographic names used on map drawings.

A number of subsystems are connected to the main system, e.g. systems for statistical calculations and drawing systems for X-Y plotters and graphic-display terminals. These subsystems are largely based on standard software, but often strongly modified, partly to make the programs more efficient and partly to adjust them to the actual computer facilities used.

Computer hardware facilities used

All data processing on ADK is done on two minicomputers (Norsk Data's NORD-10/S and ND-100) connected by a high-speed communication line. Mass storage consists of 600 Mbyte disk storage and one magnetic tape station. To the computers 20 terminals are connected and further 6 lines are reserved for (at present) 10 external users (mainly county administrations) which have dial-up access to the data and program systems on ADK.

Besides normal CRT display terminals, two graphic display terminals are used (Tektronix 4014 and 4006), 2 digitizer tables (Tektronix 4956 and Calcomp 848), 4 X-Y plotters (two Tektronix 4662, one HP-7470 and one Calcomp 1051).

The digitizer work-station consists of a digitizer, a CRT terminal and a small X-Y plotter used for initial control of the digitizing. The graphic display terminals are used for correction and modification of the digitized data.

The Parameter Data System (PDS)

The parameter data system contains the values of properties at the observation point or in a sample collected at the observation point. These properties can be coded by:
1. Numerical values representing a dimension, e.g. depth, the content of chemical components etc.
2. A numerical value, a text code, representing a name i.e. instead of recording e.g. gravel, is recorded the number '70' etc.
3. A character string, which is used when old recordings are handled, for which it is difficult to create a text coding system.

The reasons for using text code systems are:
1. The computer handling of numerical values is considerably faster than the handling of text strings. This is particularly the case when sorting data.
2. During recording of data much less space is required on recording sheets, which makes it possible to record much information on small recording sheets e.g. during field work.
3. The use of text code systems in connection with the sorting of data wanted forces the user to create systematic recording systems with unique text codes, which give better data and less errors.

Data structure

The basic principle in the PDS is to make it possible to record any feature and to be able to expand the recording systems as the need arises. A PDS project is created in the following steps (Fig. 1):
1. The properties to be recorded are defined. The text codes are defined initially, as they normally are changed or expanded during the life-time of the project.
2. The property values are placed in dataset values, which may contain:
- A numerical value.
- One or two text codes, i.e. two parameters stored in the same dataset value.
- A combined numerical value/text code dataset value.
- A string.
3. The dataset values are combined to datasets. A dataset contains a fixed number of dataset values, the same dataset values (by name) may be used in any combination in any number of datasets. Normally the first three dataset values in each dataset contain (1) an

DATA-SET VALUES

Name	Type	Used in data set
1. Locality	value	1, 2, 3
2. E-coordinate	value	1, 2, 3
3. N-coordinate	value	1, 2, 3
4. Clay %	value	1
5. Silt %	value	1
6. Sand %	value	1
7. Soil type	code	1, 3
8. CaCO3	value	1, 3
9. PH	value	1, 3
10. Map sheet	code	1, 2, 3

TEXT CODES

1 = coarse sand 2 = fine sand 3 = clayey sand etc.	1 = 1111 I NØ 2 = 1111 I SØ 3 = 1111 I NV etc.

DATA SETS

1. Soil texture - values: 1, 2, 3, 10, 4, 5, 6, 7, 8
2. Coordinates - values: 1, 2, 3, 10
3. Lime data - values: 1, 2, 3, 10, 7, 8, 9

etc.

FILES

1. Files containing data set 1 (soil texture) data
2. Files containing data set 2 (coordinate) data
3. Files containing data set 3 (lime data) data

etc.

Fig. 1. Relations between PDS data elements.

identification number for the observation (often observation point number) (2) and (3) the east and north coordinates for the observation point, values which are mandatory in most geographic data processing. A special dataset type is used for descriptive texts connected to the parameter dataset from e.g. an observation point. Any amount of descriptive text can be recorded, further the texts can be split into different types, e.g. after importance.

The defined data-set values, data-sets and text code systems are placed in a catalogue, the so-called Common Project File CPF, for the actual project.

Creating databases

A PDS database contains data decribed in a CPF. A PDS project may contain data of a single type, e.g. soil texture analyses, or data of many different types, e.g. all data collected in a specific mapping project. A PDS project database contains thus data which in most applications are to be used together, but as the system allows easy interchange of data between different projects, and combined use of data from different projects, the limits are rather wide.

A data base consists of a number of files, each file contains data recorded in a specific dataset. Each file is described in the CPF with a set of file parameters, and identified by a name. The file names and the dataset value names (numbers) are the main keys to the use of the data (Fig. 2).

When recorded data are available, the following steps are thus:
1. The data are transferred from the recording sheets to the computer. This can be done in different ways:
- By on-line writing to a file (raw data file).
- By off-line punching: the punched data are later transferred to the raw data file.
- By means of a small computer program, guiding the input and performing initial data control.
2. The raw data are either transferred directly to PDS files, or reformatted before this is done. The latter because the recording of data (on recording sheets) is done as practical as possible for the recording personnel i.e. the recording sheets often contain data belonging to more than one data-set.

Between Steps 1 and 2 is performed data control, both manually and by computer programs, and errors are corrected. Special test programs are often developed for this purpose.

User specifies file name e.g. lime1111, the CPF file parameter part is searched for this file name, when found the actual program is informed about number of datasets in file, coordinate scale, and the dataset number for the datasets in the file. The program finds the actual position in the dataset of the value to be used (e.g. dataset value no. 7). File parameters:

N datasets	Coor. scale	Data-set nr.	File name	Informa. param.
e.g. 1234	1000	3	lime1111	year, name etc.

dataset no: 3 - values:

1	2	3	10	7	8	9

value no: 7 have position no 4 in the dataset

Fig. 2. Use of file parameters.

Use of the data files

When a data file has been created, different types of standard treatment are possible:
1. Data listing:
- Standard lists, where the content of the data-set values are presented in columns. Text codes are replaced by the actual texts.
- Special lists, produced by special programs using the PDS files.
2. Map drawing: The content of the data-set values may be depicted on plotter-drawn maps in many different ways:
- As numerical values, up to 5 different values at the same point.
- As symbols, which either represent the values directly or represent a numerical interval.
- As figures (circles, squares, triangles or columns) where the side length or area represents the size of the value.
All symbols, values and figures can be drawn in different colours, sizes and thicknesses. On the maps may be drawn information from any number of files and PDS projects. Further the maps may contain LADS data (see below).
3. Standard statistical programs for all basic statistical calculations on the data are available. Further it is possible to display the data-set values as histograms, or in one, two and three component diagrams.

Database management

When the PDS data files have been created, it is possible to perform different types of database management, e.g.:
1. PDS files can be split or combined into new files, e.g. dataset values from files with two different datasets can be sorted and combined in a file with data in a third dataset.
2. Datasets can be deleted, new datasets can be inserted, and the content of the individual dataset values can be changed in different ways, e.g. as value exchange in specific datasets, or more general inserting of the content of dataset values.
The flexibility in the system allows a gradual building up of a database, and allows the creation of e.g. temporary files to be used for a short period for a specific purpose. Further it is easy to reorganize the databases if the initially chosen structure shows up to be inappropriate.

Sorting data

All utility programs contain some basic sorting facilities, which may be expanded by connecting special sorting routines to the standard programs.

Sorting on numerical values
1. Value content specifications: only datasets containing a specific value are used.
2. Value size sorting. The content of data-set values should be between two limits. This may be combined for two or three values, and may be done either as AND sorting (all values between the

limits) or OR sorting (accepted if one of the values between the limits).
In practice, these relatively few possibilities have proved to be sufficient for most purposes.

Sorting on coordinates
Sorting of data within a specific coordinate defined frame is a very important part of the system. The following possibilities exist:
1. Sorting data inside a rectangular area.
2. Sorting of data inside a circle.
3. Sorting of data inside an arbitrary coordinate polygon.

Data files

File structure
A PDS data file consists of a number of fixed length records, each with a fixed amount of datasets, depending on the number of dataset values on one dataset. Each record can therefore be regarded as a two-dimensional matrix, where a specific dataset value occurs in a fixed column. The retrieval of dataset value content is therefore very fast.

File organization
All data belonging to one dataset in a specific project may be stored in one file. But as the databases at ADK contain data from the entire country, e.g. 45 000 soil texture analyses each with 24 parameters, the data of a specific type are often split into a number of files, where each file contains data from one map sheet (scale 1 : 100 000 or 1 : 50 000). In such cases, it is possible for the user to specify type of data, and coordinates or map sheet number, and the system will then find the correct data files.

The line area data system (LADS)

Definitions

The LADS is based on a number of simple concepts, illustrated on Figure 3.
1. A LINE consists of a number of coordinate defined points, the distances between points are variable, depending on the type of data and the line curvature. A line may be:
- A boundary line between two areas, e.g. two different soil types, where the line describes two sets of properties, i.e. one set for the area to the left and one set for the area to the right of the line.
- A non-boundary line, e.g. a road or a water course where properties are connected to the line itself.
2. The following concepts are actual for lines:
- A line may be a closed line, i.e. start and end coordinates are equal.
- Lines of Type 1.2 may or may not be connected to other lines at their start or end points.

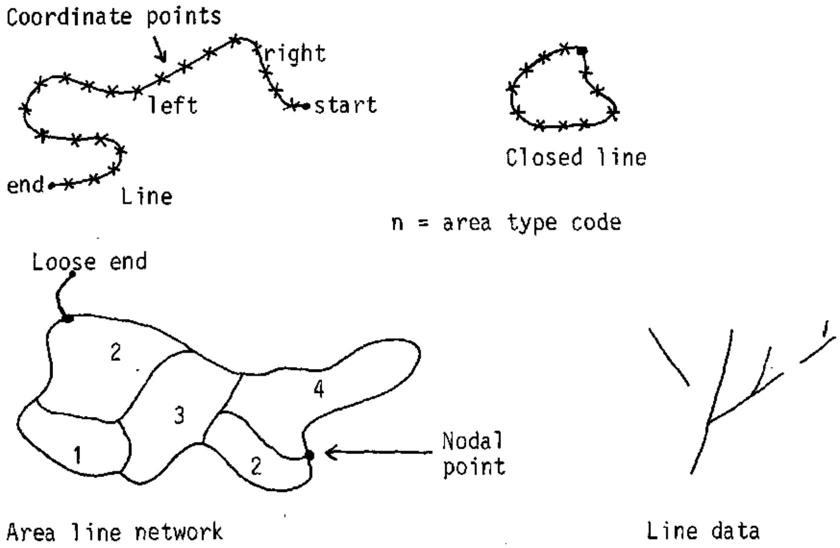


Fig. 3. LADS.line elements.

- Lines of Type 1.1, unless they are closed lines, will be connected to at least one other line at both ends. Area boundary lines therefore form a completely closed network.
- Lines meet at nodal points, line ends not connected to other lines have one or two loose ends.
- For boundary lines, properties are described in codes characterizing the area at the left and right sides of the lines.

Creating databases

The basis for generating line databases is maps, for which the following rules apply:
1. Maps in any scale be used.
2. If data from the actual map is to be used alone, a local coordinate system may be applied. But normally a general (national) coordinate system is used (at ADK the UTM system).
3. The actual map is identified in the general coordinate system by means of 4 points with known coordinates, normally the corners of the map.

Digitizing line data
The lines from the maps are transferred directly to the computer mass storage guided by a small computer program, which partly instructs the user and partly performs some basic format control of the data. Codes characterizing the lines are typed on a loose keyboard connected to a terminal, or directly on the digitizer cursor. The digitizer has a theoretical accuracy of 0.1-0.2 mm, but at ADK 0.6-1.0 mm reproduction accuracy is regarded as satisfactory,

corresponding to 15-25 m in nature for the most commonly used map scale 1 : 25 000. When a map has been digitized, the following procedure applies (fig. 4):

1. The data are transferred from digitizer coordinates to general coordinates and corrections are made for possible shrinking or expansion of the maps, with respect to their theoretical size.
2. A preliminary drawing is made on a small X-Y plotter and missing lines are digitized.

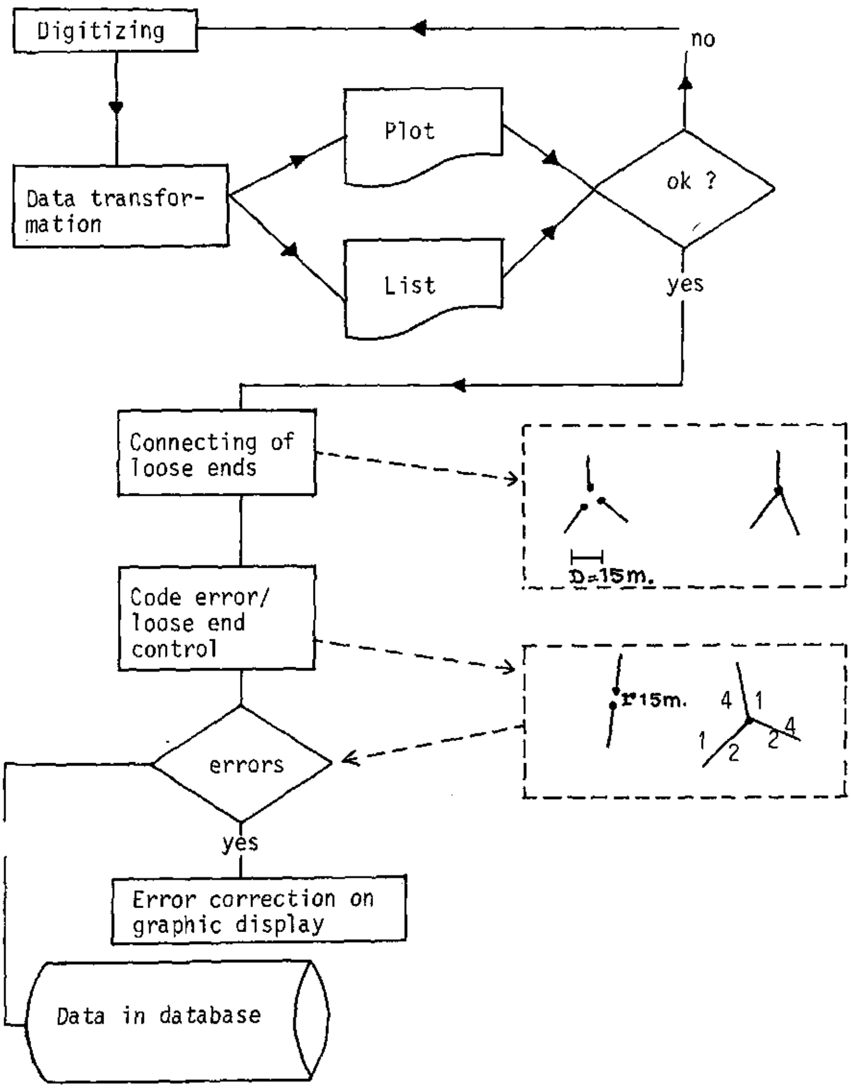


Fig. 4. Generating of LADS data.

3. A program for connecting loose line ends is run on the data.
4. Error control programs are run to detect the following errors:
- Loose ends.
- Code errors at nodal points, i.e. meeting lines should have identical codes towards the same areas.
- The outside codes for closed lines are compared with lines with boundaries towards the same area.
5. The errors are corrected by means of a graphic display terminal and an interactive program with facilities for moving lines, points, connecting lines, splitting lines and code changes.
6. Steps 4 and 5 are repeated until all errors have been corrected.
7. The map is edge-matched to the neighbouring maps; this is partly done by a computer program (Step 3). Steps 4 and 5 are also applied to the boundary zone.
8. An accurate plotter map is drawn, and the drawing is compared with the original map with respect to correspondence between original lines and digitized lines. The codes inside closed lines are also verified. The last few errors are corrected on the graphic terminal.

Updating line data

The updating of line data is done in two ways:
1. Small changes are done directly on the graphic terminal.
2. Larger changes are made in the same way as the original digitizing, i.e.
- The changed or new lines are digitized and transformed.
- The new data are added to the old data.
- The 'wrong' lines in the old data are removed on the graphic terminal, and the new lines are fitted to the remaining part of the old lines.

Database structure

Data structure

The data for a digitized line are stored in the following way:
1. As a header is stored the following information:
- A unique line number.
- Two codes defining the type of the line.
- The number of coordinate points in the line.
- Smallest and largest coordinates for points in the line.
- Default drawing colour, and default drawing type for the line (unbroken thin line, thick line, broken line etc.).
- Codes characterizing the line.
- Digitizing date or date of last modification.
2. The coordinates for the points along the line. The data stored in fixed length records, directly transferrable from mass storage to computer program. The number of coordinates is used as index to the start of the next line, and the other parameters are used to sort out the data to be used, and especially in map drawing, to specify the line type/colour to be used for presenting the relevant data type.

File organization
The data are organized in files. A file may either contain all data of a specific type, or (more common) a part of the total amount of data. The line data are in general stored in files containing data from one map sheet (scale 1 : 25 0000). If the amount of data is small, the data from the 1 : 25 000 map sheets are combined into files covering map sheets in the scale 1 : 50 000 or 1 : 100 000. By this map sheet organization of the data, the amount of data in the individual runs is reduced. As the coordinates for map corners are known, the user needs only to specify a coordinate frame and the data type to be used, the utility programs are then able to find the actual files.

Discussion
The ADK LADS differs in several ways from other systems with a similar purpose, the most important seem to be:
1. Menus are seldomly used, because it is much faster to supply the computer with the actual codes directly from a keyboard. In this connection should be mentioned, that more complex description of e.g. areas, or points along a line, is done by means of PDS data.
2. The updating of data is not done on an integrated work station where the digitizer is connected on-line to the computer and a graphical display terminal, and where the updating/change of data are guided by a menu (or codes). Such a configuration implies a relatively large investment which cannot be utilized effectively, and further the requirement to the computer capacity is much larger compared with the actual system used.

Use of LADS data

The LADS data are used for two main purposes, map drawing and area calculations, some other applications of minor importance at present are:
1. Line length calculations.
2. Finding the area type for a specific coordinate defined point.
3. Construction of lines parallel to other lines.

Map drawing
The LADS data may be depicted on maps in any scale and covering any area. The drawing of maps with PDS data is normally combined with LADS data, e.g. coastlines, administrative boundaries, water courses, to give topographic reference, or the PDS data are used to characterize the areas shown by the LADS data. On the maps, it is possible to combine any amount of data from different LADS files, and to specify by line type or codes which of the lines should be drawn. The default colours/line types may freely be changed to other colours and line types. All these facilities make it possible to produce often very complex thematic maps, where it is still possible to distinguish the different items.

Area calculations
The calculation of the areas for different types within another

area is a very important possibility in the system. A typical example is soil type distribution within a county or a drainage area. In the case of e.g. drainage areas, these areas are themselves LADS data, from which the polygon delimiting a specific drainage area can be subtracted and used as a sorting frame on other types of LADS data or on PDS data.

Map production

A procedure for producing printed multicolour maps based on data from LADS databases has been developed and tested in the production of a soil map for the island of Bornholm. The basis is partly digitized topographic information and partly digitized soil type boundaries, from original hand-drawn maps. In the construction of the soil maps, the PDS was used to produce base maps with results from the texture analyses of samples. The final line data were handled by a flatbed plotter carving out symbols and lines on scribe coat films used for creating printing plates in the printing of the multicolour map. Simpler printed maps are often produced by offset printing from plotter drawings. Multicolour print is produced by selective drawing in black of each of the colours on the printed map.

Software specifications

Language

All programs are written in ANSI-77 FORTRAN, with the extensions found in the Fortran version used on Norsk Data computers. The software is therefore theoretically easy to move to other computers with a similar Fortran compiler, although it may be rather difficult in practice.

Software structure

The nucleus of the system consists of about 75 programs/subroutines, but many of these programs occur in a number of versions, e.g. considering digitizing programs, a basic program occurs as 15 versions each handling a specific type of digitalization.

The following main groups of programs exist:
1. PDS data entry programs, used for guiding of data input and preprocessing.
2. PDS CPF maintenance program, used to create, update and print CPFs.
3. PDS standard data-file-creating programs. Read raw data into the PDS file structure and update the file parameter part of the CPF.
4. Standard PDS listing program, standard programs for statistical treatment, and a number of special utility programs producing various types of list and plotter drawings.
5. PDS data management programs for splitting or combining of files, change of the content of data sets etc.
6. LADS digitizing programs, and programs for transformation of

digitized coordinates to general coordinate system.
7. Programs for error control in LADS data.
8. Programs for updating, modification and error correction in LADS data, and various database management programs for these data.
9. Area and line-length calculation programs, other utility programs for LADS data.
10. Programs for creating polygons that can be sorted on PDS or LADS data and programs able to sort LADS data inside a polygon or split lines where they are cut by other lines.
11. Programs for handling reference coordinates for maps (map corner coordinates) handling coordinates for nearly 4 000 reference points on 5 different map sheet systems covering Denmark.
12. Map drawing program for depicting of both PDS and LADS data. Besides the main version, this program occurs in a number of versions with special layout routines added.

Totally the system at present consists of about 250 programs/subroutines. All programs are operated in a dialogue process between the user and the program. To handle all standard usage of the programs, a number of small programs are used generating the input for the actual programs, e.g. if a user needs a standard map drawing, he only needs to call such a program and give the map sheet number and the data type.

Databases handled at ADK

Since 1977, the following databases have been implemented at ADK, all databases are continuously updated. For further details about the projects to which the data belong, see ADK (1982) and Mathiesen (this volume).

PDS databases

- Soil texture/chemistry databases, from 1977-1983, 1.5 million data values.
- Water supply drillhole data, 1975-1983, 0.8 million data values.
- Forestry investigation in Denmark, 1978-1983, 2 million data values.
- Investigation of soils containing pyrite, 1982-1984, 0.5 million data values.
- Investigation of drainage conditions, 1982-1984, 2-3 million data values.
- Other projects, 1-2 million data values.

LADS databases

- Soil type database, including coastline, lakes, towns and forest areas, 1979-1982, 3.2 million coordinate sets (E,N).
- Administrative boundaries, 1981, 0.4 million coordinates.
- Watercourse, 1981-1984, 2.0-2.5 million coordinates.
- Drainage area boundaries, 1981-1983, 1 million coordinates.
- Wetland areas, 1982-1984, 2.0 million coordinates.
- Various minor line databases, 0.5 million coordinates.

At present about 80 % of the LADS data are available in databases.

Further development and problems

The present system is the result of 15 years of development, where many people have been involved in the specification of what and how the system should work. This process still continues, partly because of developments in technology, and partly because more and more people use the systems.

A main principle in the system is to be able to produce low-cost output, especially maps. The idea is to make it economically possible for the users to experiment with the output facilities from case to case, to obtain the best background for the further use of the information. The system is not intended to produce final maps, as most planning work is based on compiling information from many sources.

The actual systems contain a large number of possibilities. Therefore the problem for the user more often is to find a way to the actual output wanted, and not so much a question of modification of the system. Much of the development during recent years has therefore been concentrated on making the programs more easily usable. But although the system has become relatively easy to use, it requires much experimenting and learning for the users. It is certainly not just to press the button, and then the wanted result is delivered. The problem is to make it possible for a user to produce relatively complex output in a simple way, within a reasonable time, without creating a time-consuming dialogue between program and user.

References

ADK, 1982. Land data systems at Ministry of Agriculture Bureau of Land-data, Denmark.

Platou, S.W., 1971. An electronic data processing system for geological field and laboratory data. Rep. No 39. The geological Survey of Greenland.

Platou, S.W., 1975. Århus University System II. In: W.W. Hutchison (Ed.): Computer based systems for geological field data. p. 43-46. Canadian Geological Survey paper 74-63.

Platou, S.W., 1983a. The ADK PDS computer program system. Bureau of Land-data, Denmark (In print).

Platou, S.W., 1983b. The ADK LADS computer program system. Bureau of Land-data, Denmark (In print).

Going database in IRIS: a status report[1]

George Bluhm, John Carlis and Gordon Decker, Soil Conservation Service, US Dept. of Agriculture, P.O.Box 2890, Washington DC 20013, USA

Introduction

Increased use pressures on the USA's soil, water, and related resources have created a need for accurate information on the availability and condition of the most basic resources. This information must be timely and almost immediately available to resource managers, decision makers, and policy setters to support wise use of resources and to show the impacts of unwise use and irrevocable changes in the resource.

The Soil Conservation Service (SCS) of the United States Department of Agriculture (USDA) is making an effort to automate, integrate, and make readily accessible natural resource information as mandated in Public Law 95-192, the Soil and Water Resources Conservation Act of 1977 (RCA). The goal of this effort is to provide for the development of an information system that provides 'the right information to the right people at the right time to make the right decisions'. The system will be developed and maintained so that the user does not have to have knowledge of computer programming. This effort is being initiated by the Integrated Resources Information System (IRIS) Staff.

Resource assessment

The Soil and Water Resources Conservation Act requires that the Secretary of Agriculture appraise the nation's soil, water, and related resources and ensure that the conservation programs his department administers foster use of those resources that serves the long-term interest of the nation. The major concern that stimulated the United States Congress to require this analysis was the realization that the continued deterioration of the resource base through soil erosion, the continued losses through irrevocable changes in land use, and the limitations of water supply and quality will continue to stress our finite resources.

Figure 1 shows a conceptual flow of the information needed to guide the development of strategies for wise future use of natural resources. This process begins with basic data about soils, land use, water, climate, terrain, and vegetation that have the capacity to

1. A joint effort of the USDA-SCS, Washington, D.C., USA and the University of Minnesota, Minneapolis, through an agreement with the USDA Agricultural Research Service.

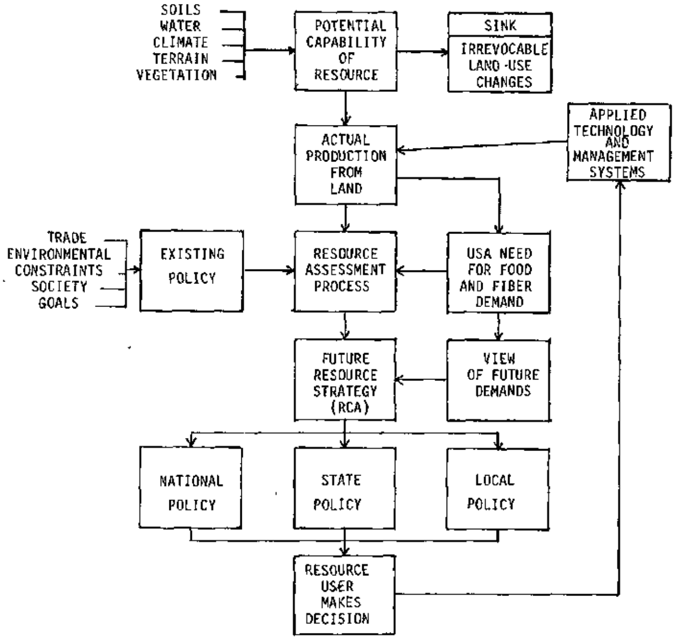


Fig. 1. Conceptual flow of the information needed for natural resource assessment and future resource strategy development.

produce food and fiber. These data indicate productive capacity of the resource base. Actual production never equals potential capacity, however, because of deterioration or inefficient use of the resource base, limitations imposed by Mother Nature in any one growing season, and the influence of incentives that affect the behavior of resource users. The goal of the Soil Conservation Service is to effectively assist land users to maintain the long-term productive capacity of the resource base while the users are wisely managing the land for short-term economic gains. To achieve that goal and meet the requirements of the Soil and Water Resources Conservation Act, SCS conducts a resource assessment and develops strategies to meet future demands.

In the RCA process, the Agency asks:
1. What are the most probable future demands on our natural resources?
2. What are and what will be natural resource problems and where are they or will they be located?
3. What will be the production capability shortfall or surplus under various strategy alternatives?
4. What are the risks and costs involved in implementation of each of the alternatives?

Answering these questions requires integration of information collected throughout the nation in a given time period (the 1982 Natural Resources Inventory) with existing resource information such as soils data that were collected over many years and were automated for a special technical use other than a national assessment.

Integrating national statistical databases that were collected at a given time with soils data that were collected over a 35-year period during which constant changes were occurring is extremely challenging. The major data problems associated with the national assessment are associated with conception and testing of schemes that will provide a more general level of aggregation than that for which the data were collected. The long-term data are site-specific for use at the county level or below, and the statistical data were collected to represent aggregations at the Major Land Resources Area (MLRA) portion of State or larger areas.

The soils information is the basic foundation upon which SCS is building other resource information systems. However, there exists some difficulty in integrating even the basic soils information. Most of the data collected over the past 35 years was descriptive in nature for a rather narrow specific technical application. Today new applications are less technical with broader application. The task of data integration includes defining the range of continuum of properties to be represented by a single integer value, assessing external conditions when that value is a good approximation of the real world, devising practical interfaces between scales of measurement, determining proper evaluation computations, and determining the physical interpretation of what the final answer means.

Data management

A normal step in the evolution of processing data in an organization is 'going database'. To some, going database simply means using a database management system (DBMS), which is a commercially available software product. To others, it means the integration of several independently developed application systems that have overlapping data requirements. To still others, it means the ability to get reports by issuing commands rather than by writing programs. Going database for SCS's IRIS staff includes all three.

ANSI has developed a three-level model of database systems that is useful in understanding our task (ANSI 1975). A part of that model is depicted in Figure 2. For each program or human user, there are separate views, or schemata, of the data stating what types of data are available. At the logical or conceptual level, there is a single view or schema stating what types of data there are. This view is essentially a union of all the user views. An important characteristic of these two levels is that they do not show how the data is stored. That function is reserved for the physical level, where there are a set of file organizations each of which positions, orders, and, in some cases, indexes a set of data.

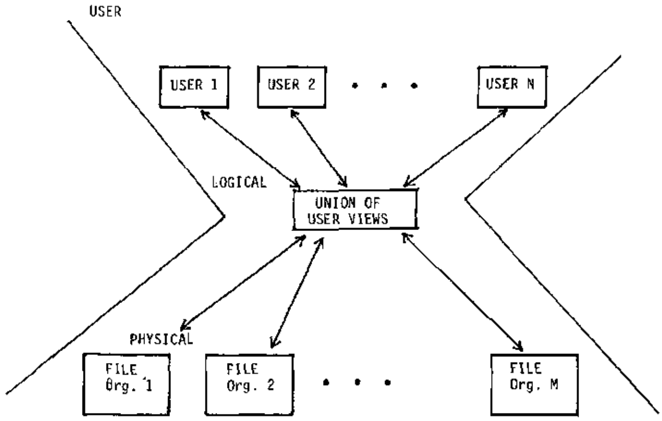


Fig. 2. Three-level model of a database system (AMSI 1975).

Going database has several advantages. Data independence is one. Data independence means that the DBMS insulates application programs from changes in how the data is stored. A second important advantage is increased speed and ease of use by two types of users - programmers don't have to write file manipulation routines and non-programmers can use a high-level command language to define, update, and retrieve data. Other, secondary, advantages are: redundancy control, inconsistency control, standards enforcement, security and integrity capabilities, and lower overall cost. See James (1977) for a discussion of these.

These benefits are achieved at some price. DBMS have overhead costs: they take up space and time in the computer, they require competent support people, and they require the intellectual effort of designing an integration system by people, in this case soil scientists, knowledgeable in the uses of the data.

To obtain the above benefits, the first step is to determine what types of data exist and how they are named and related. The IRIS staff is discovering, documenting, and communicating knowledge about data through a graphic form called a logical data structure (LDS). An LDS is simple, consisting of only five major notions: entity, attribute, relationship, relationship descriptor, and identifier. Figure 3 is a graphic representation of an LDS for a simple management application.

An entity is a generic group of objects about which information is maintained. An entity instance is a unique occurrence of an entity. In Figure 3, 'employee' is an entity and an individual employee would be an entity instance.
Entities are described by attributes and relationships with other entities. Attributes are characteristics of the entity. Attributes of an employee are the person's name, employee number, age, etc.

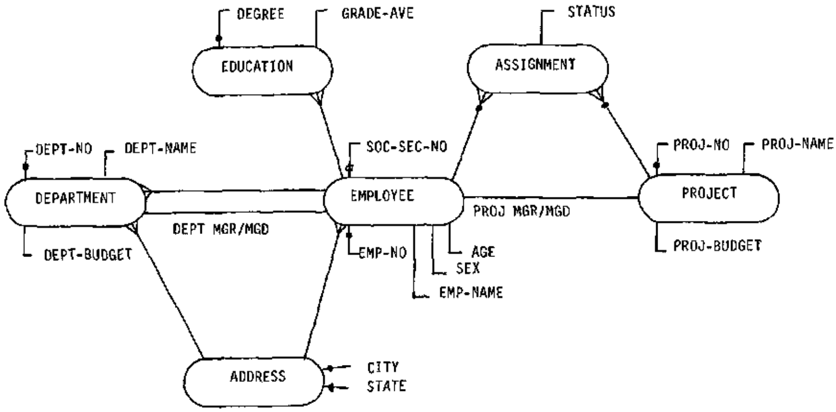


Fig. 3. Logical Data Structure (LDS) for typical employee, department, and project relationships.

Relationships are structural associations; in the LDS in Figure 3, 'employee' has a relationship with the 'department' because employees work in departments. An attribute is a single-valued descriptor of exactly one entity. A relationship is binary; two entities participate in a relationship. Each entity in a relationship both describes and is described by the other; thus, there are exactly two descriptors for each relationship. The degree of a relationship descriptor is the number of instances of the describing entity that are associated with a single instance of the described entity. Relationship descriptors may be either single-valued (degree is one) or multi-valued (degree, or average degree, if it is variable, is greater than one). At most, however, only one descriptor of a relationship may be multi-valued. In Figure 3, there are both employee-of-department and department-of-employee. Since there are many employees in a department, the degree of employees-in-departments is greater than one. Since each employee is in only one department, the degree of department-of-employee is equal to one. Multi-valued relationship descriptors are represented by 'chicken feet' in Figure 3.

The fifth LDS notion is the identifier. An identifier is a subset of attribute or relationship descriptor whose values are unique for each entity instance. Each entity has one or more identifier; the first subset is termed the primary identifier. For the example in Figure 3, the primary identifier for the entity 'employee' is the attribute 'employee number', and a secondary identifier is the attribute 'social security number'.

The intent of the LDS is to capture unambiguously the data's semantics. To do this, the LDS must be well formed, that is, attributes must be single-valued and every entity must have at least one identifier. The data must be homogeneous - every instance of an attribute must have the same meaning. An example of non-homogeneity is the FIPS code, in which a value may refer to a state or (almost) a continent. Given that the LDS is well formed, the proper way to judge its quality is its fidelity of meaning. Does it express the user community's view of the world? Does it use meaningful words? A poor choice of words means that the LDS won't communicate. A poor structure won't capture the meaning and therefore can't store useful data.

An LDS can be developed by examining existing documents and programs and by interviewing users. Documents exist for each application system but are of varying quality. Names of data and encoding schemes differ for different applications. Program source coding may be the only accurate repository of how data gets transformed. For manual systems and especially when integrating independently developed applications, interviews with real people (here, soil scientists) are absolutely necessary for uncovering subtleties in the data and for recognizing and resolving conflicts in the different definitions and user applications of the data.

The soils LDS

The Logical Data Structure (LDS) approach to database management system design has been used for several years on management-type data. The SCS is applying the LDS approach to database management system design on natural resource data. Figure 4 illustrates how LDS data types relate to soil data. 'Soil series' is an entity. 'Soil series name' is one attribute of the entity 'soil series'. Relationships between entities are identified by a line connecting the entities; and each relationship has two relationship descriptors. There are both (many) soil series-of-taxonomic classification and taxonomic classification-of (one) soil series. There are many soil series in one soil taxonomic classification but only one taxonomic classification of a series. Each entity has one or more unique subsets of identifiers. The entity 'soil series' is identified by both the soil series name and the taxonomic classification. The entity 'taxonomic classification' is identified by a code as the primary identifier and by a name as the secondary identifier.

The LDS for the primary SCS soil survey data is being developed as the first step in 'going database' with SCS's natural resource data The documentation of this data knowledge has been acquired through the use of existing documents and through meetings and workshops with soil scientists, economists, and other resource scientists who are familiar with natural resource data. There have been many discussions and disagreements over data definitions, data relationships, and attribute identifiers. This process has forced all data collectors and users to look at the whole and come to an agreement on definitions and relationships. This is essential for all DBMS's and for obtaining consistent results in any evaluation and analysis of the resources.

Figure 5 illustrates an LDS for some of the soils data that have

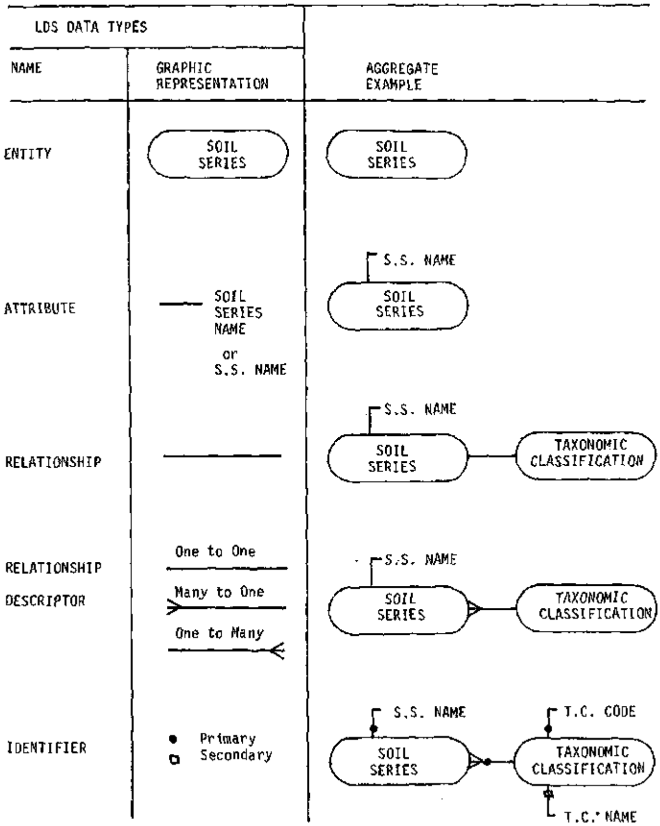


Fig. 4. Logical data structure (LDS) notions as they apply to soils data.

been documented thus far. The first component of the LDS developed was for the soil series-soil taxonomy (USDA 1975) relationships. This relationship is simple compared to relationships for the other soil data components. For example, starting with the soil series, there are many soil series in one family, many families in one soil subgroup, and so on through each higher level in the classification system. There is only one of each soil family attributes, such as taxonomy particle size, for each soil family; but there are many soil families that have the same taxonomy particle size.

The second component of the LDS shown in Figure 5 is the geographic relationships of the soil map unit. There are many soil map units in one soil survey area, but each soil map unit is in only one soil survey area. Note that there is a county soil survey area combination entity between the county and the soil survey area entities. This is required because one county may contain more than

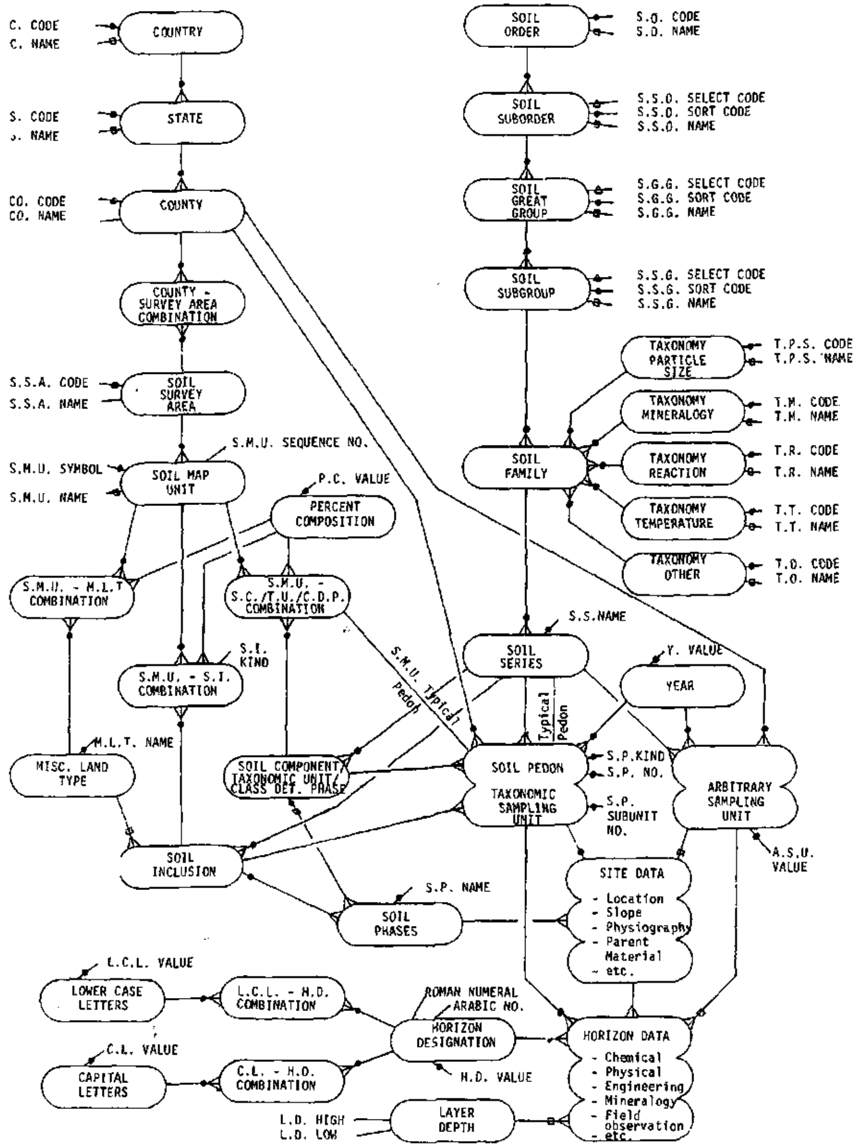


Fig. 5. Abbreviated Logical Data Structure for some soils data relationships. Included are the Soil Series - Soil Taxonomy relationships, the soil map unit-geographic location relationships, the soil map unit-component relationships, and the soil pedon data relationships.

one soil survey area, and one soil survey area may contain more than one county. A direct relationship between county and survey area would represent a many-to-many relationship, which is not allowed in an LDS.

The third component of the soils LDS is the soil map unit to soil composition relationships. The reality here is that one soil map unit may have no or many soil components, no or many miscellaneous land types, or no or many soil inclusions. One soil component, one miscellaneous land type, or one soil inclusion may occur in more than one soil map unit. This requires us to put in three combination (assignment) entities so that we can document that the appropriate soil component et al. is in combination with the appropriate soil map unit. Note that the soil component and the soil inclusion entities are identified by both soil series entity and the soil phase entity.

The last component shown in Figure 5 is the sampling unit-site and horizon data relationships. This illustrates that one soil may have many soil pedons, which we call taxonomic sampling units, and may have many arbitrary sampling units. One sampling unit has one set of site data such as location and parent material, and one site may have many sets of horizon data such as field-observed data (color, texture, structure), chemical and physical data. Note that the horizon data primary identifier is horizon designation if the data are from taxonomic sampling units and the horizon data secondary modifier is layer depth if the data are from arbitrary sampling units.

Also note the one-to-one labeled relationships between the soil pedon entity and the soil series and the soil map unit soil component/taxonomic unit/class determining phase combination entities. This relationship reflects the typical pedon that represents the official series description of the soil series and the typical pedon for the soil components in the soil map unit.

The LDS for all the soil attributes associated with the sampling unit site and horizon data have been developed. This structure accounts for most of the kinds of soil attributes collected. We will continue to expand the LDS until all the soil data are documented. The complete LDS will include all the other resource data, such as data for yields and land-use, that have been collected on sites where the kind of soil has been identified.

The completed LDS will enable us to determine the kind of DBMS that would be the most efficient and cost effective for managing our natural resource data. The step following the selection of the proper DBMS will be the development of the physical data structure (PDS).

IRIS has already developed an interactive retrieval system for the soils-5 interpretation data. To use the system, the user must know either the soil series name or the soils interpretation record number. Another aspect of the same system allows the user to enter soil descriptive properties and a state location, if desired, and the system will provide first the number and then a listing of all the soils series in the designated area that fit the conditions requested. The data are updated monthly with 500 to 800 changes or additions to the files at a cost of approximately $200,000 per year in systems support and $300,000 per year in survey scientists support.

After the implementation of the total integrated soils system in 1985, the SCS will be able to retrieve all textual and quantitative properties of a given soil phase; perform searches on specific properties of soils; specific classification or taxonomy of soils; and, if the soils map has been digitized, go from location to soils to interpretation. The long-range goal of the effort is to provide electronic graphic display of soils at the field survey level to assist in both survey operations and land-use decisions.

References

ANSI (American National Standards Institute), 1975. Committee 3/ SPARC/study group interim report. Association for computing machinery, Special Interest Group on Management of Data. Bulletin 7, no. 2.

James, M., 1977. Computer data base organization. Ed. 2. Prentice Hall.

USDA (United States Department of Agriculture), 1975. Soil taxonomy: a basic system of soil classification for making and interpreting soil surveys. Soil Conservation Service, United States Department of Agriculture Handbook 436, 754 p., ills.

From POSEIDON to NEPTUNE - Software for environmental surveys

A.M. Aubry et al.[1]

After the major change in host computers that occurred after our meeting in Druzhba, the project to convert the system from Fortran 5 to Fortran 4 appeared inappropriate. Therefore, a more sophisticated system was designed, flow-charted, programmed in PL/1 on an Amdahl 470/V7 CNRS-CIRCE computer. Consequently, the new software was applied to factual data from two divergent terminologies.

The experiment currently under way is with descriptors from the second terminology (on exhibit: Recherche d'un langage transdisciplinaire pour l'étude du milieu naturel. Travaux et documents Orstom, No 91, 143 pp., 1978). No doubt, the system can as well be implemented on any terminology, as expressly stated through the *acronym NEPTUNE Nouvel Ensemble de Programmes pour Terminologie* Usuelle Non-formatée à l'Entrée. Users preferring the first acronym need not change, as it is spelt out PL/1 Oversimplified System Ensuring Immediate Databank Operation in Natural Language or Programmes Organisés de Saisie, d'Edition et d'Interrogation pour Descriptions Ordinairement Non-ordonnées.

Data input

By data input, character strings only are keyed in. Character strings as defined by the system are of 3 types: enumerated strings, equality strings and ordinary strings.

(1) *Enumerated strings* are 1 to 56 characters in length, with digits, capital letters, blank spaces, and a few EBCDIC special characters % * / : ; , ' - _ $ and a terminating period. The remaining characters are system oriented. At least one character is not a digit. Blank spaces before the string are not read. The name 'enumerated string' reminds the user that it is to be found in the

1. A.M. Aubry, A. Beaudou, Y. Boulvert, E. Braudeau, Y. Chatelin, J. Collinet, J.F. Dupon, J.C. Filleron*, M. Fromaget, A. Garcia Gómez, A. Giey**, Ph. Horent*, B. Koli Bi*, R. Poss, J.F. Richard, G. Riou and R. Van den Driessche;
Office de la Recherche Scientifique et Technique Outre-Mer, research unit 'Etudes intégrées et cartographie des milieux naturels et humanisés par les méthodes de la typologie et de l'informatique', 70 route d'Aulnay, F 93140 Bondy, France;
Université Nationale de Côte d'Ivoire (*), Abidjan;
Sociologiciels, 20 rue Ste Croix Bretonnerie, 75004, Paris (**), France

nomenclature, which, let us stress the point, can be updated instantaneously. By the way, the nomenclature is split into a (sampling) unit section, whose descriptors are used once in a record, and a subunit section, whose descriptors are used repeatedly. Within each section, the descriptors are aggregated into variables. All data are of the ordinal or nominal type. Names of the variables are not used by data input, only by search.

(2) *Equality strings* contain a simple equals sign = followed by a signed integer in the range -32767 to 32767 (optional positive sign) and a period. Left of the = character is a string 1 to 56 characters long (same as above) or the abbreviated 1 to 4 characters synonym explicitly stated in the nomenclature (at least one character not a digit, no blanks within the synonym). Both strings may be spelt the same way when this seems preferable. Data are of the interval type in the great majority of cases.

(3) *Ordinary strings* are introduced by a plus sign + and terminated upon a period. All EBCDIC characters are used here, except . and @. Such an ordinary string may be unique, 1 to 400 characters in length, or split into unequal strings dispersed within the unit description. To avoid lengthy records, the current system excludes ordinary strings from the subunits as there are 20 subunits in a unit. Comments are the most frequent type of ordinary string. However, an uncertain, biased, or even highly significant descriptor may enter as an ordinary string with the appropriate remark, duplicating the enumerated or equality string. Also when there are subunits in excess, the additional ones may enter as ordinary strings.

Just a hint about keywords; 2 logical keywords are mandatory: the names given to unit and subunit in the form of equality strings, e.g. H=7.L=4. --- .L=2. ---. L=1. -----. Before input, a visual check of key presence and accuracy is advocated. Subunits always follow their unit and may be entered in any order. Moreover, units need not follow the order of their integer identification number.

Validation

Validation is omnipresent as storage occurs solely upon validation of the strings found in the nomenclature. By validation, 3 situations occur:
1. The *enumerated string* is read and then located in the nomenclature by scanning, if necessary, the whole list; if not found, it is printed and annotated by an interrogation mark.
2. For an *equality string*, both string and integer are subject to control: string must be found and integer must lie within the specified range; here again, verification is concluded by printing the input string either plus a period or plus an interrogation mark.
3. An *ordinary string* cannot be validated; total length, however, must comply with limits; with overflow, an arrobas @ sign replaces characters in excess that had nevertheless printed and were truncated.

When two or more descriptors from the same variable enter the same subunit or unit, the last one only is validated and stored, although all are output on the lineprinter. What occurs when a period is not keyed in? The resulting concatenated string is rejected, printed and annotated, even if both descriptors are valid!

Nomenclature

Within the nomenclature, strings for the unit description are presented first; the second part, under the title subunit, covers all strings available repeatedly. Enumerated strings are not presented alphabetically, but in a logical order suited for later editing (see below). They are aggregated under a fixed number of variables. The variables too receive names: the usual name and an abbreviated name, both available by search, never by data input (the two names need not be different). Such a variable is followed by a list of strings, each one being matched with an integer value. To a given integer can be associated more than one string: synonyms, abbreviations, codes ... , all available upon data input and upon search. Such a list is permanently open to receive new strings and its content may be reordered (first string is privileged by editing). A descriptor never has the same spelling as a variable; the difference may be in a single character. For equality strings, the nomenclature contains only the two strings and the anticipated minimum and maximum, the narrower the range the better the validation. Obviously, all strings of the nomenclature must differ to permit recognition and non-redundancy is checked.

The limits set on a nomenclature file and the current status are as follows.

	Maximum	No 6	No 7
within-unit			
number of variables	151	151	151
number of strings	1000	811	416
number of characters	32752	8965	6319
within-subunit			
number of variables	18	16	16
number of strings	1000	730	333
number of characters	32752	4814	3119

Nomenclatures No 6 and No 7 contain 16 subunit variables (*) and 2 dummy variables for later use. When upper limits are reached, it may prove advantageous to build the system with several homologous nomenclatures, as explained in Orstom Init. Doc. Tech. No 25, 1974. Number 7 contains numerous redundancies and has, nevertheless, been stored, on the same disk, by a modified storage programme lacking the 'non-redundancy' constraint and that for editing purposes only.

Storage

The host computer is an AMDAHL 470/V7 linked to an IBM 370/168 under MVS+JES3 at CNRS-CIRCE, on the Orsay campus. At least 320 terminals cover the country. Furthermore, a second computer, an IBM 3033/UO8, at CNRS-CNUSC, on the Montpellier campus, has been introduced in the same network. One of the ORSTOM remote batch terminals is a CII-HB MINI6 computer with a dozen implemented keyboard+video consoles. For the files, indexed data are the key to organization on a resident 3350 disk. Integers from the nomenclature are stored in binary fixed, the ordinary strings are concatenated and stored in the same record as character string. The fixed length record has following structure:

- a non-embedded key 5 bytes;
- the 151 integers from unit 302 bytes,
- the character string 400 bytes,
- the 20 subunits of 18 integers each 720 bytes;

with a total record length of 1422 bytes, or 711 half-words of 16 bits. Missing values are stored as an integer specified in the nomenclature (-32768). Integers are ordered as in the nomenclature.

Update

Update may be local or global. The latest read descriptor of a given variable is the one to get storage. We take advantage of this by LOCAL update. For a mispelling, for instance, unit and subunit (when relevant) are identified and the descriptor typed correctly. Local update permits also erasure of a given descriptor by a single # keystroke (key £ on ASCII console). By GLOBAL update, an externally stored record (on floppy) is modified on the video, read, and stored, erasing completely the previous record bearing the same number.

Edit

To obtain lineprinter output in plain language of individual records or of a series of records, we give their identification numbers, e.g. 1801, 1609 A 1622. Editing a complete retrieved subfile is by giving its own number.

Search

All strings currently in the nomenclature are available for search purposes. Ordinary strings cannot intervene. Enumerated strings and equality strings are used simultaneously, in any number, in short or long spelling and in any order. When appropriate, a subscript (in parenthesis) is added to the string in order to restrict search within the corresponding subunits. There is no need to learn a search command language: the terminology is familiar to the user, and the number of symbols (system-oriented and nomenclature) has been reduced.

Applicability of the operators ¬ & | on operands type A > B, A < B, A = B, where A and B are strings or integers, under normal priority rules, or in parenthesis, offers as wide a spectrum of search capabilities. On local keyboards, press ∧ key for logical negation ¬

(not), & for logical conjunction (and), and ! for logical disjunction | (or). Parentheses are available to modify priority. The negation is used as follows: ¬ string = integer ¬ string = string (with = > or <). All integers, even out of range, can be used. No periods are allowed within the search formula, only one final period. The search strings are validated and the logic of the syntax is checked.

The search formula is converted to reverse-polish form:

```
if      A ,   B ,   C , . . .   are operands
and     a ,   b ,   c , . . .   are operators
a ( ( A b ( B c C ) ) b ( ( B c C ) d D ) )
is written
A B C c b B C c D d b a
```

Search is in batch mode and output is on the lineprinter, or on console CRT via a local disk, as an ordered list of record numbers. These records may be retrieved in totality as a subfile bearing the same physical and logical characteristics as the main file. This enables the user to follow the search on the subfile and this, of course, may be repeated as NESTED SEARCH. Lineprinter output in plain language of the retrieved records may be requested at any step of the nested search.

Studies on soil pH and $CaCO_3$ distribution in soil profiles based on a soil data bank

Henryk Kern, Institute of Soil Science and Plant Cultivation, Izabelli 5/8, PL-24100, Puławy, Poland

The first paper on the information system of agricultural environment work out at the Institute of Soil Science and Plant Cultivation was presented at the second meeting of the ISSS working group in Varna 1977 (Kern, 1977). The present paper describes the system and presents the first system-aided results on pH and $CaCO_3$ distribution in the soil profiles.

The system concerns mainly soils of farming lands and some factors of the natural environment, such as agroclimate, relief and water relations. The aim of the system is to provide data within the area of administrative or natural units needed by various organizations or individuals for better use of agricultural land. The system was designed as an open system to which more data could be included as needed. The smallest unit for reference data is a village or state farm. It is fully possible to collect the data for any area of the administrative unit-commune or voivodship (province) as well as natural ones: physical and geographical, the catchment area and so on - according to need.

Sources of data for the system:
- legends of large-scale soil agricultural maps;
- results of chemical and physical analyses and descriptions of soil profiles;
- results of routine analyses of arable horizons (pH, P, K, Mg);
- some climatic factors (e.g. precipitation, temperature);
- some data on soil contamination.

The database of the system comprises all data collected (and put in writing on 14 forms) during soil survey, in the Voivodships of Gdańsk and Opole 1965-1980 (where the system was made available to users and introduced).

The database consists of five files:
1. The Dictionary (codes and names of administrative and natural units).
2. The main (all legend descriptions of soil units, soil complexes)
3. The profile (results of analyses and descriptions of profiles).
4. The climate (average data on temperature and precipitation for many years).
5. The arable horizon (results of investigations by agrochemical stations on pH, P, K, Mg).

The system provides printouts of data of practical use for agri-

culture, e.g.:
- calculations of soils useful for definite kinds of plants;
- surface determination: of heavy soils for tillage, of calcareous or acid soils (in whole profile), of soils with water surplus or water deficiency;
- correlation coefficients for a given unit (e.g. textural groups with $CaCO_3$);
- assessment of soil reaction and needs of available nutrients;
- suitability classes (for soil, relief, climate and water condition).

This system uses the computer ODRA-1300 (ICL compatible) or other with memory > 45 K, card reader and punch, and line printer. The software is written in COBOL (Babiarz et al., 1978).

The database on magnetic tapes and software are held in the Computer Department of the Institute in Puławy. The copies of database and software will be available on line for voivodship authorities of Gdańsk and Opole. The institute will hold all documents for the voivodships (in the future for all 49).

Studies on soil pH and distribution of $CaCO_3$ in soil profiles have been conducted on the 'old profile' database (about 48 000 analyses) collected for the country in 1965-1975 (Kern & Pietraś, 1981).

At the beginning, soil profiles were distinguished by occurrence of $CaCO_3$ within textural classes. It was found that most samples containing carbonates originated from soils rich in clay particles (grain size below 0.02 mm) (Fig. 1). Then, $CaCO_3$ content of soils was related to percentage sand, silt, clay and colloidal fraction (Fig. 2). The following relationships were found:
1. Negative correlation with sand fraction
2. No relations with silt fraction
3. Positive correlation with clay and colloidal fractions.

At the second stage of studies, differences in $CaCO_3$ distribution in the soil profiles were investigated in relation to age of soils. A large group of soils developed from loams (sandy and medium loams) with $CaCO_3$ in profile on two glacial landscapes, were chosen for investigation. The selected files, containing 1673 soil profiles, were divided into two parts comprising two glacial landscapes in Poland: 1. The Baltic Glaciation (Würm); 2. The Central Polish Glaciation (Riss). While comparing the same soil units in the two glacial areas, statistical differences in the mean content of carbonates were found. Soils developed from boulder loams of the Baltic glaciation contained on average more $CaCO_3$ than those of the Central Polish Glaciation (Fig. 3, Table 1). Differences were found in the occurrence of $CaCO_3$ in soil profiles of three identical soil units (Fig. 4a), and in average $CaCO_3$ values in soils of both glacial areas (Fig. 4b). The younger soil units of the Baltic Glaciation are richer in $CaCO_3$ and contain this component higher in the soil profile (A_1, (B) horizons). By contrast, older soil units of the Central Polish Glaciation are poorer in $CaCO_3$

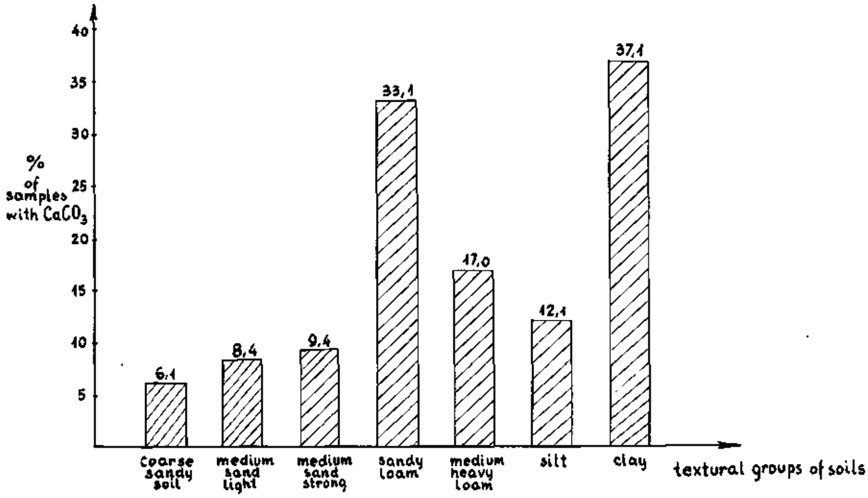


Fig. 1. Percent of samples with $CaCO_3$ in textural groups of soils.

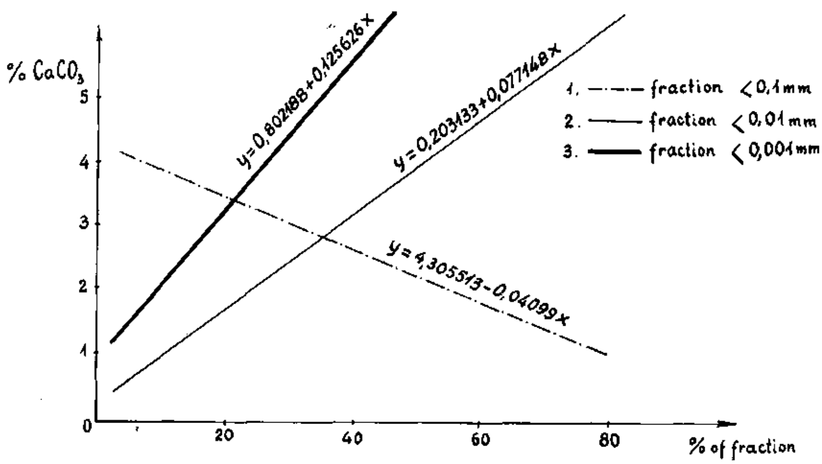


Fig. 2. Content of $CaCO_3$ in soils according to % share of fractions.

Table 1. Content of $CaCO_3$ in soil units within compared areas.

No	Soil unit	% Of samples with $CaCO_3$ (on each 100 samples)		Mean content of $CaCO_3$ in soil profiles		Least significant differences
		Baltic glaciation area	Central Polish glaciation area	Baltic glaciation area	Central Polish glaciation area	
1	Medium sand on loam	17.0	5.6	5.60	3.35	3.33
2	Sandy loam	31.3	28.0	6.87	3.00	2.74
3	Sandy loam on medium heavy loam	28.2	27.5	6.94	2.28	2.17
4	Medium heavy loam	57.0	39.7	7.26	2.49	2.18
Average for all soil units		33.3	25.2	5.57	2.93	1.11

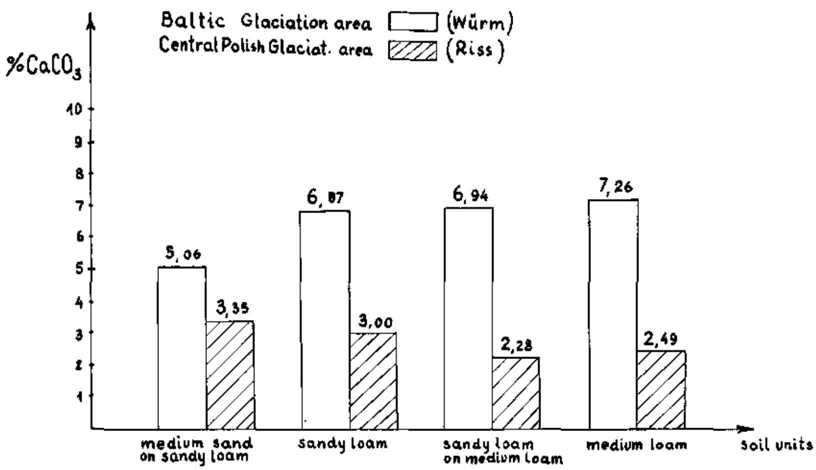


Fig. 3. Content of $CaCO_3$ in the soil units.

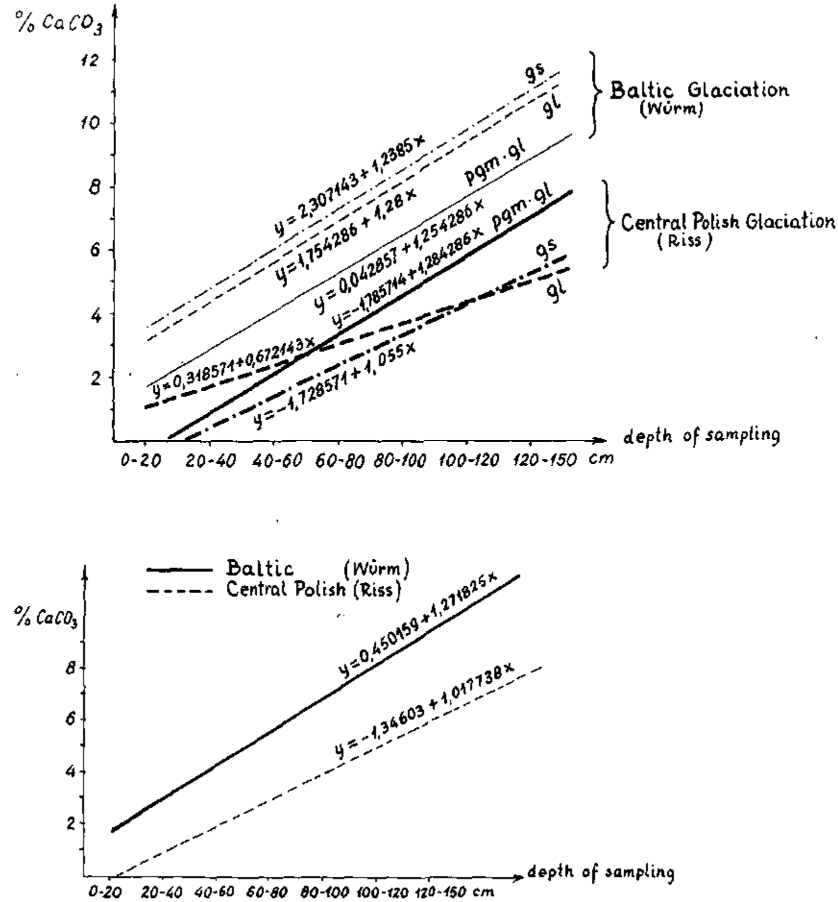


Fig. 4. Distribution of $CaCO_3$ in soil profiles developed from sandy loams and medium heavy loams localized on two glaciation areas: Würm and Riss. Fig. 4a (above). Three identical soil units. Fig. 4b. Average values.

and contain it at a lower horizon (C or D).

These differences indicate stronger leaching processes, and a deeper decalcification of soils of the Central Polish Glaciation area.

The results provide adequate information to predict the lime requirement of particular soil units, developed from loams. The database also allowed initiation of a special kind of soil map - the soil decalcification map, based on the soil agricultural map. The problem of soil acidification and distribution of acid soils within Poland have been discussed in other papers (Kern, 1982, 1984).

References

Babiarz, R., H. Kern & M. Kosiński, 1978. Technical project of information system in Institute of Soil Science and Plant Cultivation at Puławy. ZETO-Lublin.

Kern, H., 1977. Progress in application of computer methods for the agricultural environment systems in Poland. In: Developments in Soil Information Systems. (Dr Alexander M. Sadovski and Dr Stein W. Bie Eds). Proceedings of the Second Meeting of the ISSS Working Group on Soil Information Systems in Varna, Sofia, Bulgaria. 30.V - 4.VI. p. 74-77. Pudoc, Wageningen.

Kern, H., 1982. pH value and $CaCO_3$ content in soil profiles of Poland. Papers of the Meeting on Soil Liming in Poland - Nov. 1982.

Kern, H., 1984. Some problems of geochemical evolution of Polish soils. Geographia Polonica Nr 50 (in press).

Kern, H. & B. Pietraś, 1981. Decarbonization degree of Polish Soils on basis of Bigleb Information System. Roczniki Gleboznawcze T.XXII, Nr 3. Warszawa (in English).

Soil classification in Denmark and its adjustment in relation to land-use planning

F. Duus Mathiesen, Ministry of Agriculture, Bureau of Land Data (ADK), Enghavevej 2, DK-7100, Vejle, Denmark

Abstract

In the period 1975-1980, a comprehensive soil classification of Denmark was carried out by the Bureau of Land Data (ADK) and presented as multicoloured soil maps in the scale of 1:50 000 for the whole of Denmark. Today these soil maps are widely used in physical and agricultural planning in Denmark. To adjust the soil classification to the needs of agricultural productivity and environmental planning, various research programmes have been initiated at ADK. These include examination of soil profiles in the path of gas pipelines across Denmark, mapping of soil types containing pyrite, survey of drainage conditions related to soil types and topography, crop identification with Landsat data, and forest inventory. These programmes and a programme on nitrate leaching are briefly described, and a model is presented for calculation of irrigation requirements of agricultural areas.

Introduction

In Denmark, agricultural land constitutes about 70 % of the total area, equivalent to 2.9 million hectares. Since 1938, the agricultural land has been reduced by more than 300 000 hectares. Part of the former agricultural land - especially marginal land - has been afforested.

The notable reduction in agricultural land is one of the main reasons for the incorporation of provisions into the agricultural legislation, aiming at protection of the diminishing tillable land resources in Denmark. Also, public interest has grown for balanced development of open land resources through active planning. This planning has to take social, economic and ecological aspects into account, and to provide directives for land-use with respect to urban development, traffic systems and other infrastructure, agriculture, forest and strip mining areas, conservation, and areas for recreation.

Reasonable land-use planning requires an effective and flexible land-data system. To meet the demand for information at different administrative levels, the Ministry of Agriculture's Bureau of Land Data (ADK) has developed an extensive database and an electronic data-processing system, with special emphasis on soil, forest, water and raw materials.

The system has been working since 1975, and is today widely used in public planning and administration. This paper provides a short description of the ADK data-processing system and some of the programmes being carried out by ADK.

ADK data-processing system

The ADK data-processing system is based on a system designed for thematic mapping. The system has been further developed at ADK to handle the Danish Soil Classification, and a number of other subjects.

The data system consists of functions treating data elements separately or in combination. The various data groups are used for graphical presentations, statistics, and for calculations, for instance of areas related to selected parameters such as soil types and catchment areas (Fig. 1).

Research programmes undertaken by ADK

Research programmes undertaken by ADK are listed in Figure 2.

Danish soil classification

A national soil classification of Denmark was initiated in 1975 with the purpose of gathering information on Danish soil resources (Fig. 3, Fig. 4).

The classification was completed in 1980 with 400 multicoloured soil maps (scale 1 : 50 000 covering the whole of Denmark. These maps have subsequently been digitized in order to provide facilities for the production of computer-drawn soil maps in different scales and with different combinations of parameters. Areas of soil types, for instance, can be calculated by the computer system for administrative regions or catchment areas, within any chosen section of the maps.

All basic data from the soil classification are stored in an easily retrievable database which is continually updated with new soil data.

Danish forest inventory

Since 1979, a Danish forest inventory programme has been carried out by ADK in co-operation with the State Forestry Enterprise and the forest organizations. The purpose of the inventory is:
- to establish a basic information system, which can easily be brought up to date,
- to improve administration of forest legislation,
- to complete a national survey of small-enterprise forestry.

The programme includes survey of forest areas meeting requirements of directives given by FAO/ECE. Boundary lines are digitized in the

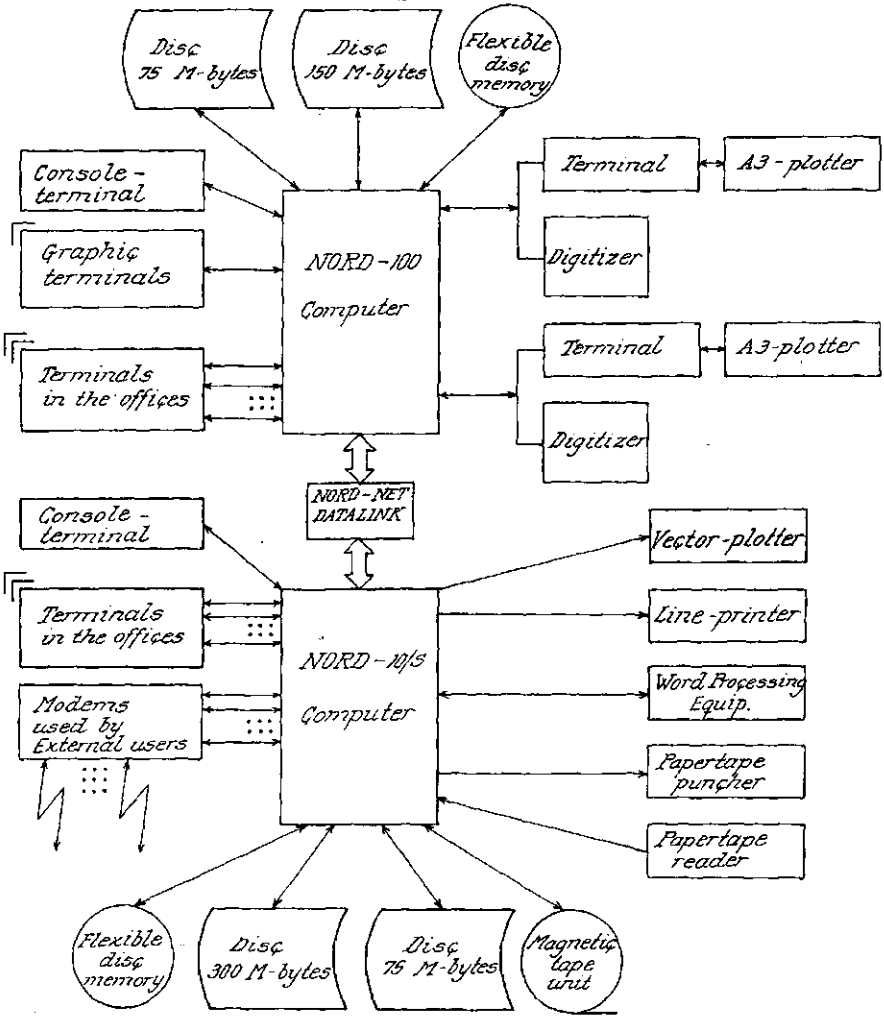


Fig. 1. Diagram of electronic data processing equipment at Ministry of Agriculture's Bureau of Land Data (ADK).

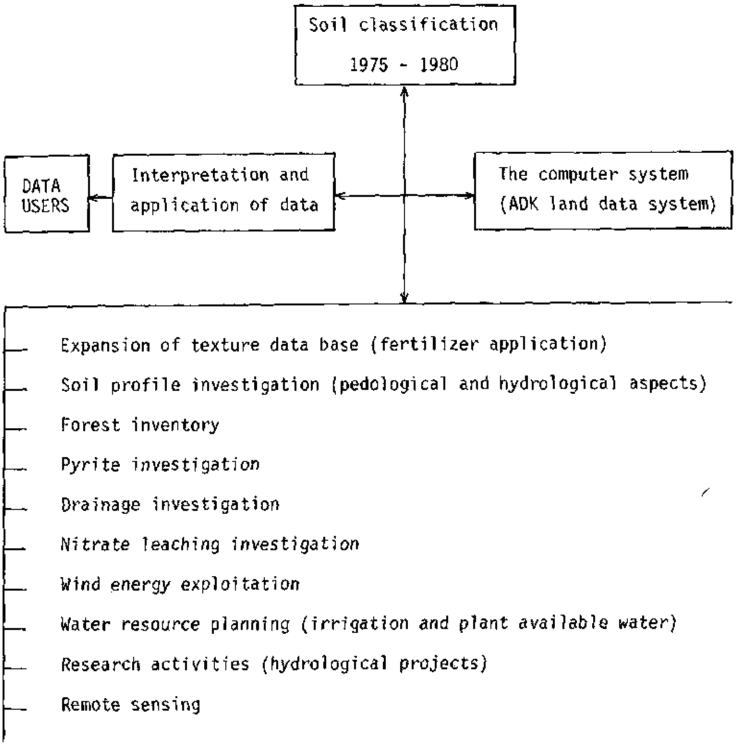


Fig. 2. The Danish soil classification and subprojects undertaken from 1980 to 1985 by ADK.

UTM system in order to facilitate map-drawing and calculation of areas (Fig. 5). The various forest areas are described with information on tree species, ages and wood production.

Survey of soil profiles

The construction of pipelines across Denmark from Danish off-shore gas fields represents a unique possibility of studying soil profiles in various geological deposits. The investigations are carried out by ADK in co-operation with the Government Laboratory for Soil and Crop Research, the Geographical Institute at the University of Copenhagen, the Chemical Institute at the Royal Veterinary and Agricultural University and the Geological Survey of Denmark. The investigations are financed by the Danish Natural Science Research Council, the Agricultural and Veterinary Science Research Council and the Ministry of Agriculture.

The ADK project comprises a pedological classification of soil profiles at intervals of 25 m in the pipeline trench, which is about 2 m deep. Detailed pedological studies are carried out with

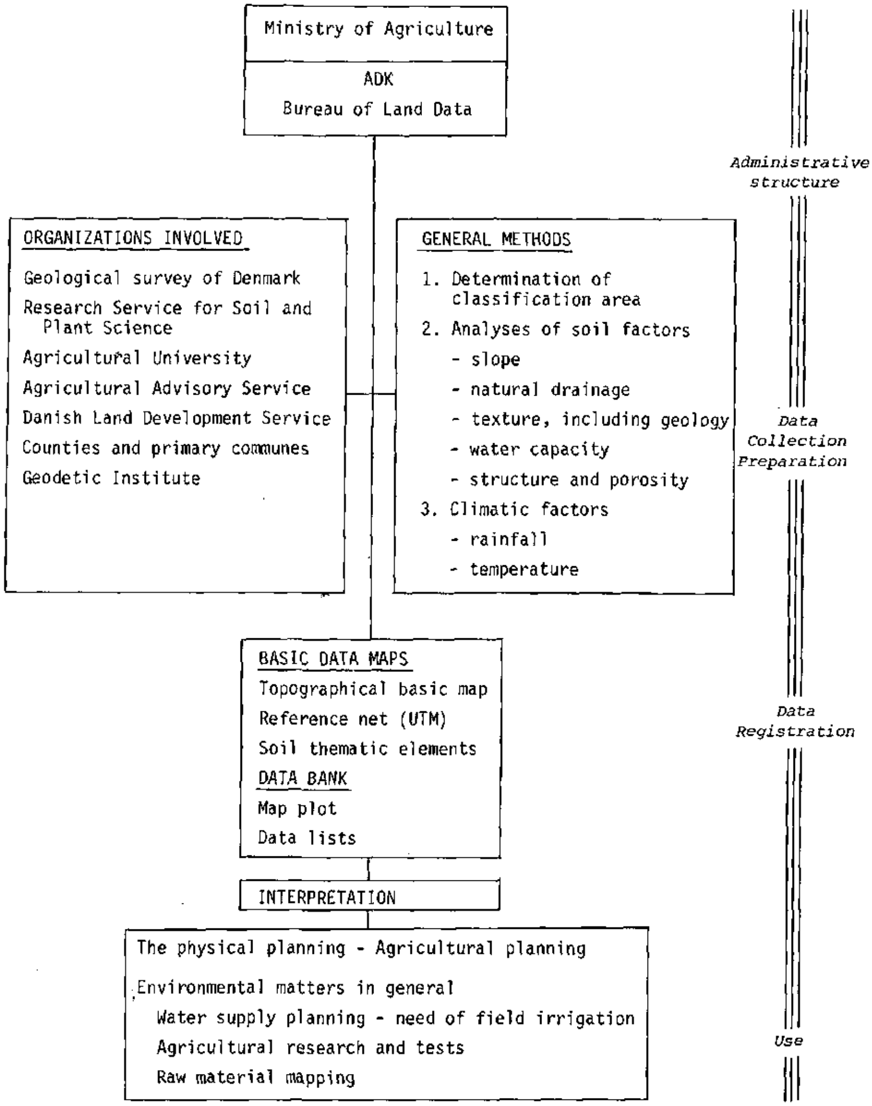
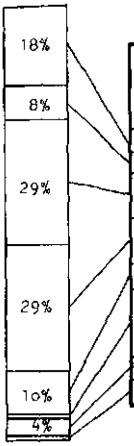


Fig. 3. Structure and organization of the Danish soil classification project 1975-1980.



Map Colour Code	SOIL TYPE	JB nr.	Percentage by weight: Clay < 2 µm	Silt 2-20 µm	Fine Sand 20-200 µm	Total Sand 20-2000 µm	Humus 58.7% C
1	Coarse Sand	1	0-5	0-20	0-50	75-100	≤ 10
2	Fine Sand	2			50-100		
3	Clayey Sand	3	5-10	0-25	0-40	65-95	
		4			40-95		
4	Sandy Clay	5	10-15	0-30	0-40	55-90	
		6			40-90		
5	Clay	7	15-25	0-35		40-85	
6	Heavy Clay or Silt	8	25-45	0-45		10-75	
		9	45-100	0-50		0-55	
		10	0-50	20-100		0-80	
7	Organic Soils	11					> 10
8	Atypic Soils	12					

Fig. 4. Definition of soil types in the Danish soil classification system, and the average distribution (%) in Denmark of the various types. The Map Colour Code is a reference to the different colours representing the dominant soil types in the map series: Soil Classification of Denmark. The JB nr. is a subspecification of the soil type, related to the data files of texture analysis.

3 profiles every 1000 metres along the trench. Samples are taken from every horizon in each profile. The samples are analysed for texture, organic content and various chemical and physical parameters, and all descriptions and data are stored in the computer database.

The results have scientific use as well as practical for evaluating the productivity of the soils. The project will extend knowledge of soil development in various geological deposits. The detailed profile studies will add new information to the soil maps already published by ADK. Water-retention measurements from the profiles will provide the basic data for detailed evaluation of the requirement for irrigation. The investigations of soil profiles will furthermore contribute to a better understanding of the problems related to the leaching of nutrients (e.g. nitrate) from the soil.

Investigation of soils containing pyrite

The water quality of Danish water courses can be severely reduced by strong acidity and a high concentration of dissolved iron. A detailed mapping programme of poorly drained soils containing pyrite has recently been initiated by ADK in cooperation with the Ministry of Environment.

Survey is based on records of pyrite concentration in soil profiles. For the programme, more than 7000 profiles have been described in

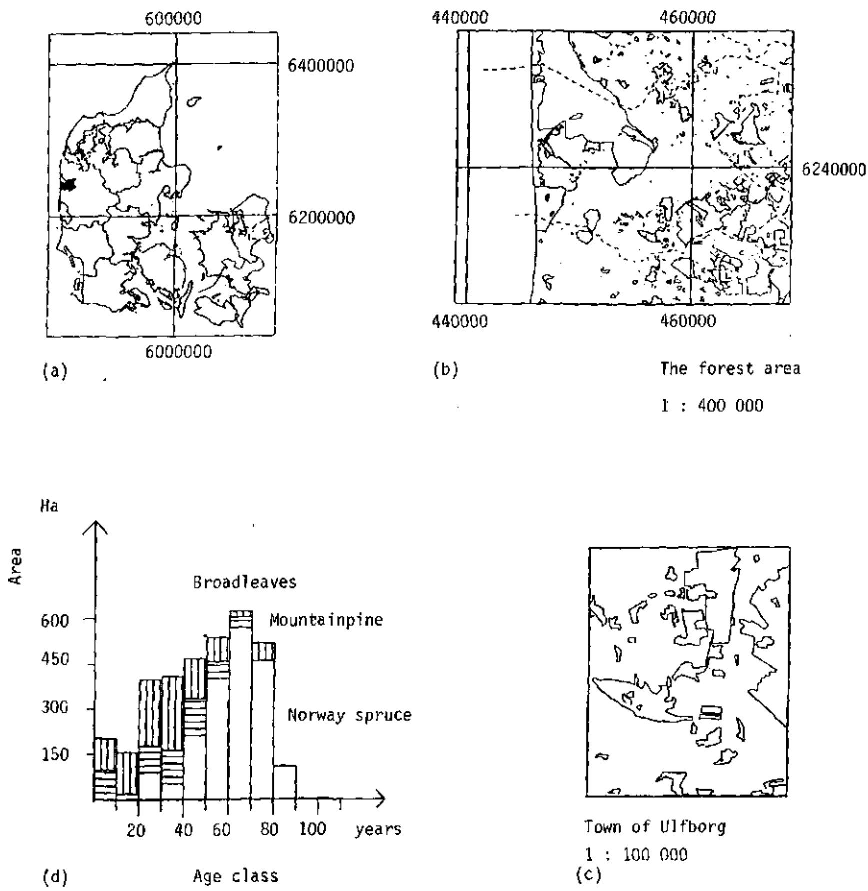


Fig. 5. The Danish national forest inventory. Examples of plots: (a) The location of Ulfborg-Vemb commune in Denmark, (b) distribution of the forest area in the commune, (c) forest areas around Ulfborg town, and (d) the area, age-class and tree-species distribution in the commune.

detail. All data are stored in files at ADK, where computer-based maps are produced depicting areas with high pyrite concentrations.

The results will be used in planning for public action against infiltration of water courses by acid water enriched by dissolved iron, a problem which in certain regions may be exacerbated by necessary drainage of agricultural areas.

Survey of drainage conditions

The recently initiated survey of drainage in Denmark gathers information on types of soils and landscapes with a special need for drainage. It provides data for balanced evaluation of the requirement for drainage, and for assessment of environmental impact of drainage on watercourses. Results from the drainage investigation is presented for the agricultural sector and planning authorities as computer-drawn maps.

The surveys of drainage and pyrite together provide a basis for evaluation of potential acid sulphate conditions in connection with drainage (Table 1). The results should be a useful tool in preparation of future directives on drainage.

Table 1. Example of calculation of the size of wetland and potential acid sulphate areas in a number of Danish catchment areas. Case 1 and 2 represents different drainage conditions as a function of the difference in watertable.

Catchment area code	Length of river	Catchment area	Wetland area		Area of potential acid sulphate	
			Case 1	Case 2	Case 1	Case 2
	(KM)	(HA)	(HA)	(HA)	(HA)	(HA)
4047	20.2	15000	1200	300	300	250
4201	8.1	9000	200	90	50	50
4109	10.7	12000	500	300	100	80

Survey of nitrate leaching to the sea

In co-operation with the Research Institute for Plant Science and the Ministry of Environment, ADK has surveyed leaching of nitrate and phosphorus into enclosed Danish coastal waters (the Baltic Sea and the Kattegat).

The survey encompasses more than 70 % of the Danish land area (Fig. 6). The area was divided into 24 subareas, based on 123 catchments and the distribution of soil types and land-uses were calculated for each subarea (Table 2). In order to estimate the amount of nitrate leaching from agricultural land to the sea, leaching values as a function of soil types were established (Table 3). On average in the period 1975-1981, about 40 % of the nitrate discharged into coastal waters was derived from agricultural land.

Table 2. Investigation of nitrate leaching. Soil type distribution and land-use (ha).

LANDBRUGSMINISTERIET - AREALDATAKONTORET
ENGHAVEVEJ 2
7100 VEJLE

DATO.: 1982-07-01

ALLE VANDLØBSOPLANDENE MED AFSTRØMNING TIL ØSTERSØEN OG KATTEGAT (123 OPLANDE IALT)
AREALTYPEFORDELING FOR OPLANDENE.: 801 TIL.: 6502

BEMÆRK: JORDTYPEBIDRAG FRA NORDJYLLANDS AMTSKOMMUNE ER BESTEMT UDFRA PRØVESTEDSFORDELINGEN I 1975 - KLASSIFICERINGEN.

AREALERNE AFRUNDET TIL NÆRMESTE.: 100 HEKTAR

TYPE 1	(GROVSANDET)	=	298800	HEKTAR	9.8 %
TYPE 2	(FINSANDET)	=	232500	HEKTAR	7.6 %
TYPE 3	(LERBLAND.SAND)	=	751600	HEKTAR	24.7 %
TYPE 4	(SANDBLAND.LER)	=	783700	HEKTAR	25.7 %
TYPE 5	(LERJORD)	=	189700	HEKTAR	6.2 %
TYPE 6	(SVÆR LERJORD)	=	10100	HEKTAR	0.3 %
TYPE 7	(HUMUS)	=	148900	HEKTAR	4.9 %
TYPE 8	(SPECIEL JORDTY)	=	7500	HEKTAR	0.2 %
TYPE 9	(BYZONE)	=	210300	HEKTAR	6.9 %
TYPE 16	(SOMMERHUSOMR.)	=	3200	HEKTAR	0.1 %
TYPE 17	(LANDZO.LOKPLAN)	=	300	HEKTAR	0.0 %
TYPE 18	(LANDZ.LOKPL.SH)	=	0	HEKTAR	0.0 %
TYPE 10	(FERSKE VANDE)	=	35200	HEKTAR	1.2 %
TYPE 12	(REST OMRÅDER)	=	33700	HEKTAR	1.1 %
TYPE 13	(SKOV OMRÅDER)	=	323400	HEKTAR	10.6 %
TYPE 14	(SKOV I BYZONE)	=	1100	HEKTAR	0.0 %
TYPE 19	(SKOV I SOMHSOM)	=	0	HEKTAR	0.0 %
TYPE 15	(IKKE KLASSIFIC)	=	14800	HEKTAR	0.5 %
TYPE 20	(PRÆKVARTÆR)	=	0	HEKTAR	0.0 %
	TOTALT	=	3044800	HEKTAR	100.0 %

BYZONE (9+14+16+17+18+19)	=	214900	HEKTAR	7.1 %
SKOV IALT (13+14+19)	=	324500	HEKTAR	10.7 %
SKOV I BYZONE (14+19)	=	1100	HEKTAR	0.0 %

FORDELING MELLEM JORDTYPER 1 - 8

TYPE 1	(GROVSANDET)	=	298800	HEKTAR	12.3 %
TYPE 2	(FINSANDET)	=	232500	HEKTAR	9.6 %
TYPE 3	(LERBLAND.SAND)	=	751600	HEKTAR	31.0 %
TYPE 4	(SANDBLAND.LER)	=	783700	HEKTAR	32.3 %
TYPE 5	(LERJORD)	=	189700	HEKTAR	7.8 %
TYPE 6	(SVÆR LERJORD)	=	10100	HEKTAR	0.4 %
TYPE 7	(HUMUS)	=	148900	HEKTAR	6.1 %
TYPE 8	(SPECIEL JORDTY)	=	7500	HEKTAR	0.3 %
	TOTALT	=	2422800	HEKTAR	100.0 %

Table 3. Preliminary assessment of areic mass rate of nitrate nitrogen leached as a function of soil type.

	Soil types				Forest
	1, 2	3	4, 5, 6	7, 8	
Nitrate N($kg \cdot ha^{-1} \cdot year^{-1}$)	8	14	19	20	5

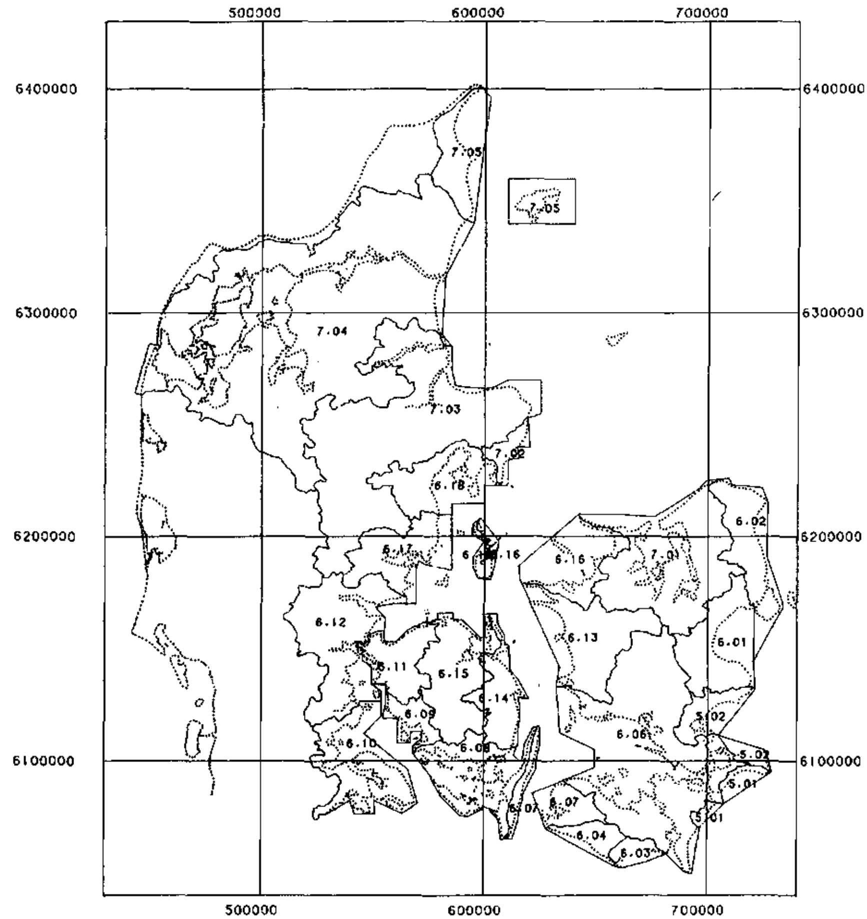


Fig. 6. Subareas used in the nitrate investigation.

Crop identification based on digital multitemporal analysis of Landsat data

Land-use data derived from satellite images may in future be considerable help in prediction of agricultural yields. ADK and the Geographical Institute at the University of Copenhagen have made a study supported by the Danish Natural Science Research Council on the use of satellite images for crop identification.

Six Landsat images were analysed and the results related to ground data from a selected area (eastern Funen). The preliminary results show that coarse identification can be made for some Danish crops by multitemporal analysis of data from 3 to 5 carefully chosen periods within the growing season.

However the spatial resolution of the present Landsat data (80 m) is insufficient for detailed mapping in Denmark. New satellite projects should provide data of higher resolution, allowing more detailed mapping.

Model for calculation of the potential need for irrigation

Agricultural planning includes estimates on future irrigation needs in Denmark and is based on data from the soil classification.

At ADK, data for regional agricultural planning is produced, including calculations of the demand for irrigation within different regions such as counties. These calculations are based on a model established by ADK. Digitized soil maps based on 43 000 texture records are used in the model in combination with data about water retention, root development and climatic conditions. The available water content in soil layers, the root zone capacity and effective rooting depth are defined and used as basic parameters in the model, which provides data on the potential need of water for irrigation, and thus data for the planning and administration of Danish water resources.

Future organization of Danish Land-data information systems

In order to coordinate the work of authorities using land data, a committee has been established in Denmark under the supervision of the Ministry of Environment. The Ministry of Finance, the Ministry of Environment, the Ministry of Agriculture and the Ministry of Housing are represented in the committee together with county and municipal organizations.

The guidelines for the work of the committee can be described as follows:
- to analyse the requirements for computerized handling of data on land, natural resources and the environment
- to secure interaction between the various Danish data-processing centres
- to investigate ways of standardizing data-collection methods.

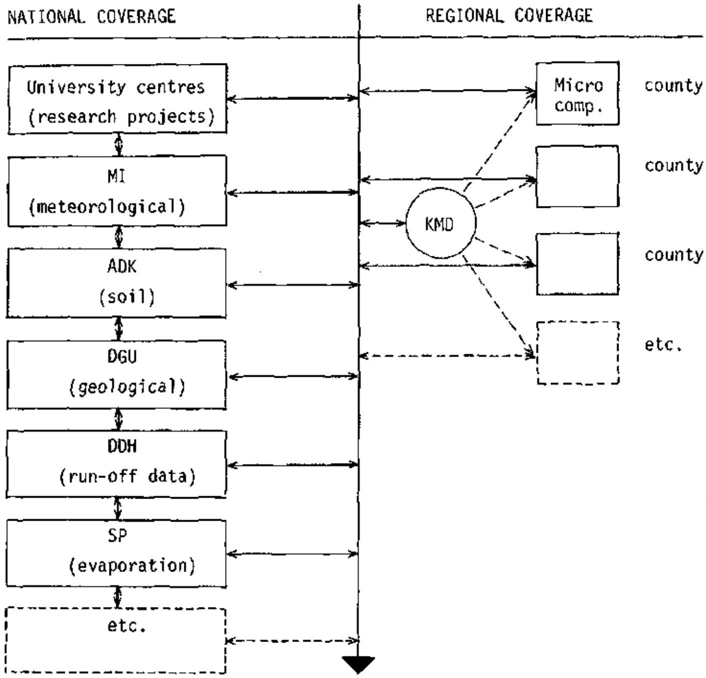


Fig. 7. A possible future organization of land data related computer systems in Denmark.

In future, emphasis will be placed on an effective interaction between various Danish data centres (Fig. 7). This will provide an essential and improved basis for the physical and administrative planning of Danish land resources.

Selected literature

Commission of the European communities (EUR 6875), 1980. Land resource evaluation. Report from a seminar in the EEC programme of co-ordination of research on land use and rural resources held at An Foras Taluntais, Johnstown Castle Research Centre, Wexford, Ireland, 1978.

Gregersen, A. & H. Knudsen, 1981. Normalvaerdier for vandingsbehov, afstrømning og nettovandbehov ved forskellig rodzone-kapacitet. Tidsskrift for Planteavl, beretning nr. S.1537.

Landbrugsministeriet, 1976. Den danske Jordklassificering. Teknisk redegørelse. (The Danish soil classification, a technical report; in Danish). 91 pp.

Madsen, H.B., 1980. Fastlaeggelse af okkerpotentielle områder i Himmerland. (English summary: Mapping of areas with a high potential of ochre in Himmerland). ADK beretning A-01. 42 pp.

Madsen, H.B., 1979. Jordbundskartering og bonitering, thesis. (English summary). Folia geographica Danica X(5). pp. 203.

Madsen, H.B., 1982. Himmerlands jordbundsforhold. (English summary). Ministry of Agriculture, Bureau of Land Data. pp. 361.

Mathiessen, Fl. Duus, 1978. Soil classification in Denmark. Its results and applicability. EEC seminar on land resource evaluation. Wexford, Ireland, 7th-9th November, 1979, 26 pp.

Mathiesen, Fl. Duus, 1982. Soil classification as a data source in relation to physical planning. Nato seminar, Land and its uses, actual and potential, an environmental appraisal, Edinburgh, 1982.

Mathiesen, Fl. Duus, S.W. Platou & P.M. Plum, 1979. Datasystem for det åbne land. (A data system for physical planning; in Danish). Ugeskrift for Jordbrug, 124 årg., hfr. 29. 5 pp.

Ministry of Agriculrure, Bureau of Land Data, 1982. Land data systems, paper prepared for an EEC workshop at ISPRA.

Platou, S.W., 1971. An electronic data processing system for geological field and laboratory data. The E.D.P. system Agto. Rep. no 39. The geological Survey of Greenland.

Platou, S.W., 1975. Aarhus University System II. In: W.W. Hutchison (Ed.): Compter-based systems for geological field data, p. 43-46.

Platou, S.W., 1982. The ADK computer program system for handling of parameter data. Landbrugsministeriet, Arealdata-kontoret (in print).

A computerized Danish pedological soil classification system

Hendrik Breuning Madsen, Ministry of Agriculture, Bureau of Land Data (ADK), Enghaveveij 2, DK-7100, Vejkw, Denmark

Introduction

From 1975 to 1980, a soil classification of farmland in Denmark was carried out. This classification entailed soil sampling at about 35 000 sites from 0 to 20 cm depth, and at selected sites samples from 35 to 55 cm depth. Texture, organic matter and content of calcium carbonate were determined in all samples. On the basis of these analyses, soil maps on a scale 1 : 50 000 were constructed, mainly showing the texture in 0-20 cm depth (Mathiesen, 1978). All soil maps and data from sampling sites were digitized and stored in a computer at the Ministry of Agriculture's Bureau of Land Data (ADK).

The computerized data have been widely used in physical and agricultural planning, e.g. calculation of water supply for irrigation and calculation of the actual and potential need of drainage within larger regions. For these purposes, a need has arisen for more detailed information about the different soil types, especially drainage classes, water retention, root development, variation in texture and the pedogenesis within the uppermost 3 metres of the profile.

Pedological investigations at ADK

The establishment of the main gas pipeline from the North Sea gasfields across Denmark in 1981-1984 presents a unique opportunity for studying soil profiles and the variation in soil properties within the mapping units on the earlier soil maps. So pedological investigations along the main gas pipeline were carried out by the Bureau of Land Data in cooperation with the Government Laboratory for Soil and Crop Research, the Geographical Institute at the University of Copenhagen, and the Chemical Institute at the Royal Veterinary and Agricultural University, Copenhagen.

The investigation comprises the following soil studies along the main gas pipeline, of which about 300 km are being investigated in detail:

- about 800 profile descriptions including sampling for analyses.
- about 8 000 classifications of soil profiles in the 2-metre-deep trench.

The 800 profile descriptions were carried out at the edge of the working area in order to get undisturbed samples. The profiles are described according to an ADK manual (Madsen, 1982) which in many respects is similar to the FAO Guidelines for soil profile description. The horizons are described according to type, thickness, texture, organic matter, colour, mottles, structure, consistence, pans, cementations, roots, stones, gley, coarse pores, coatings and boundary between horizons. All attributes are assigned values and are stored in numerical form in a computer.

From all profiles, samples were taken according to horizon sequence and analysed for texture, organic matter, pH and content of calcium carbonate. From selected profiles, dithionite-citrate and pyrophosphate iron, and aluminium, cation-exchange capacity, exchangeable bases, total N and P, clay mineralogy, water retention and root densities were estimated. All results from the analyses and the position of the profiles are stored in a computer, which makes it possible to combine the profile data with the soil map data.

A pedological soil classification system for Danish soils

In order to classify the soil profiles along the trench and for later computer handling it has been necessary to develop a Danish soil classification system (DPJ) based on easily detectable stable characteristics in the soil, so the system can be used directly in the field by a trained surveyor.

The classification system is hierarchic with four levels: order, group, series, and phase. At the first three levels, the system is based on 14 diagnostic horizons and 21 profile characteristics (Madsen, 1983), while at phase level criteria are texture in 0-40 cm and in 80-120 cm, together with pH at 1 metre depth. The soil types have been named partly in accordance with other systems and partly given new Danish designations.

As all classifications are stored in a computer the soil names can be transformed into numerical form. Table 1 shows the nomenclature at order, group, and series level and its transformation into numerical form. The structure of the system allows search for single characteristics within the profile, e.g. placic horizons, peaty top layers and gley features within a certain depth.

The soils are classified according to parent material, pedogenesis and soil properties in relation to cultivation. It proved useful to group together soils with similar texture and organic matter, soils with similar horizon sequence, e.g. those having layers impeding roots and water and soils with the same drainage conditions.

The system describes mainly the horizon sequence in the upper 120 cm of the soil and includes all diagnostic horizons or profile characteristics within the depth. Certain diagnostic horizons may lie deeper, e.g. in soils with thick A1 horizons.

As the pedogenetic and edaphic significance of certain characteristics depends on depth, the upper 120 cm of the profile are subdivided into 3 sections of 40 cm, so it is possible to weigh characteristics according to the interval where they start or finish. Thus the following combination of names may occur in a podzolized soil with groundwater gley at different depths.

Groundwater gley starting at	0-40 cm	= Podzoltypigley
	40-80 cm	= Gleytypipodzol
	80-120 cm	= gleyey Typipodzol

The properties used in the classification must be stable with reference to gley, texture, and horizon sequence, in order to make the classification valid for a longer period. So base saturation and pH in the topsoil were not included in the classification, since liming is widely practised in Denmark.

Definition of soils at order level

The profiles are classed into 12 orders based on the presence or absence of diagnostic horizons. A total of 14 diagnostic horizons have been defined. These describe the thickness of the A1 horizon, the presence of solid rock or very calcareous material, and the presence of certain B horizons, very bleached or sandy horizons, gleyey horizons within certain depths, or thick peat layers. The 12 orders can be divided into 3 large groups: deep non-hydromorphic soils, deep hydromorphic soils, and shallow soils. Table 1 lists the 12 orders with a brief definition. Furthermore the dominant soil type is listed according to FAO/UNESCO (1974).

Definition of soils at group level

The 12 orders are subdivided at group level on the basis of horizon sequence, thickness of A1 horizon, depth to solid rock or calcareous material, colour, texture, state of decomposition of organic matter, and depth to horizon with groundwater gley, pseudogley, or stagnogley. At group level, the name is constructed by writing up to 3 prefixes to the left of the order name starting with first group name nearest the order name. The names are written together and they always start with a capital letter, e.g. Gleytypipodzol. Typi is the first group name and gley the second group name.

Most of the deep non-hydromorphic soils may be given up to three characteristics at group level, namely
- horizon sequence (1st group designation)
- groundwater gley, stagnogley or pseudogley (2nd group designation)
- double-profile development in the same soil (3rd group designation).

When characterizing the horizon sequence, much care must be taken to ensure it is typical for the order, that the profile is homogeneous in texture, and that the colour fulfils certain conditions. If groundwater gley, or stagnogley starts at 40-80 cm depth, or if pseudogley is found within the upper 80 cm of the profile, this is described at second group level to the left of the horizon sequence

Table 1. Outline of the Danish soil classification system.

DRY SOILS					
SERIES	GROUP 3 GROUP LEVEL	 2 GROUP LEVEL	 1 GROUP LEVEL	ORDER	
01 kolluvial Al: 40-80 cm thick 02 humusfattig Al: <1 % organic matter 03 humøs Al: 7-20 % organic matter 04 histic peatlayer 10-40 cm thick 05 entic Al <10 cm thick 06 mor morlayer >10 cm thick 07 gleyey 08 stagnogleyey 09 pseudogleyey 10 bleget pale subsoil 11 degraderet degradated Bt or Bs 12 fragi fragipan		11 Gley 12 Stagnogley 13 Pseudogley	1 Typi 22 Bleg 6 Kalk 7 Rendzin 8 Ranker	01 rajord "Regosol"	Al <2 cm no diagnostic B horizon
		11 Gley 12 Stagnogley 13 Pseudogley	1 Typi 2 Blandings 6 Kalk 7 Rendzin 8 Ranker	02 blegsol "Arenosol"	2 cm < Al < 80 cm below Al pale colours
		11 Gley 12 Stagnogley 13 Pseudogley	1 Typi 2 Struktur 3 Blandings 6 Kalk 7 Rendzin 8 Ranker	03 brunsol "Arenosol"	2 cm < Al < 80 cm sandy brown soil without Bh or Bs horizon
		11 Gley 12 Stagnogley 13 Pseudogley	1 Typi 2 Struktur 3 Blandings 6 Kalk 7 Rendzin 8 Ranker	04 brunjord "Cambisol"	2 cm < Al < 80 cm clayey brown soil without Bt horizon
	23 Brunsol 24 Brunjord 26 Podzol	11 Gley 12 Stagnogley 13 Pseudogley	1 Typi 2 Blandings 3 Band 4 Degra 6 Kalk 7 Rendzin 8 Ranker	05 Lessive "Luvisol" "Acrisol"	Al < 80 cm soil having Bt horizon
	25 Lessive	11 Gley 12 Stagnogley 13 Pseudogley	1 Typi 2 Humus 3 Sesqui 4 Brun 5 Initial 7 Rendzin 8 Ranker	06 podzol "Podzol"	Al < 80 cm soil having Bs or Bh horizon
		11 Gley 12 Stagnogley 13 Pseudogley	1 Typi 26 Podzol 25 Lessive 6 Kalk 7 Rendzin 8 Ranker	07 kolluvial jord	Al > 80 cm

13 placic placic horizon 14 hardnet cemented layer 15 natric high content of sodium 16 kalkholdig weakly calcareous layers 17 rendzin 18 ranker 31-39 name of order+like 41-52 top+name of order 61-72 sub+name of order	6 Kalk		1 Typi 26 Podzol 25 Lessive 7 Rendzin 8 Ranker	08 stagnogley "Gleysol"	stagnogley within the uppermost 40 cm	WET SOILS
	6 Kalk	22 Bleg 25 Lessive 26 Podzol	1 Typi 2 Vad 3 Brun 4 Kolluvial 7 Rendzin 8 Ranker	09 gley "Gleysol"	ground water gley within the uppermost 40 cm	WET SOILS
	6 Kalk	7 Rendzin 8 Ranker 25 Lessive 26 Podzol	1 Fibri 2 Hemi 3 Sapri 4 Kopro 5 Blandings	10 histosol "Histosol"	peatlayer ($>$20 % organic matter) more than 40 cm thick	WET SOILS
	23 Brun 22 Bleg 25 Lessive 26 Podzol	11 Gley	1 Typi 2 Sedi 3 Blandings	11 rendzina "Rendzina"	$>$30 % $CaCO_3$ normally within the uppermost 40 cm	SHALLOW SOILS
		11 Gley	1 Typi 2 Kolluvial 3 Litho 21 Rajords 22 Bleg 23 Brun 26 Podzol 25 Lessive	12 ranker "Lithosol" "Rankers"	lime free rock normally within the uppermost 40 cm	SHALLOW SOILS

Group level
1 - 5: special horizon sequences for the orders.
6 - 8: kalk = 5-30 % $CaCO_3$, rendzin = $>$30 % $CaCO_3$, ranker = limefree rock. All three horizons begin within the uppermost 80 cm of the profile.
11 - 13: Groundwater gley or stagnogley beginning between 40-80 cm depth or pseudogley beginning between 0-80 cm depth.
21 - 26: Presence of other diagnostic horizons which might have qualified the soil at order level.

Series level
7 - 9: Groundwater gley, stagnogley and pseudogley beginning between 80-120 cm depth.
10 - 16: Horizon beginning within the uppermost 120 cm of the soil.
17 - 18: Rendzin = $>$30 % $CaCO_3$, ranker = limefree rock. The two horizons begin between 80-120 cm depth.
31 - 39: Weakly expressed pedological developments without formation of diagnostic horizon.
41 - 72: Buried soils. The deepest soil profile beginning in the uppermost 60 cm qualifies the profile at order level. The other profile is described at series level with top or sub in front of the name of order.

characteristic. If the upper limit of a gley characteristic lies above or below those limits, this will characterize the profile at order and series level, respectively, and no description of drainage conditions at group level will therefore be made. The third characteristic at group level is used to describe double-profile development in the same parent material, i.e. two pedological processes with formation of diagnostic B horizons within the same profile, e.g. a podzol developed in a lessive A2 horizon. As soils with this kind of profile development are rare, three characteristics at group level are seldom.

For deep hydromorphic soils too, up to three designations at group level are used, except for stagnogley soils, through lack of field observations. For gley soils, the first group name describes depth to a strongly reduced horizon, thickness and humus content of the A1 horizon, presence of rock and very calcareous material, the second group name is used for description of other distinct pedological developments in the profile, e.g. podzolization or clay migration, and the third group name is used to separate calcareous non-rendzina-like gley soils. As to the histosols, the first group name is used to describe the decomposition state of organic matter, the second to describe pedogenesis in the profile below the peat layer, and the presence of rock near the surface, or highly calcareous material, and the third to separate calcareous non-rendzina-like histosols.

In shallow soils, the rendzinas have three group names, whereas rankers so far have only two group names. For rendzinas, the first group name describes the calcareous horizon, the second hydromorphic characteristics within the upper 80 cm, the third soil layers above the calcareous horizon. For rankers, the first group name describes thickness of the sediment cover and the pedological development above the rock, and the second describes distinct hydromorphic characteristics within the profile.

Definition of soils at series level

When subdividing the soils at series level, profile characteristics are used that either designate the presence of certain horizons within given depths, or the chemical and physical state of different horizons. If these characteristics are absent, the naming terminates at group level. If there are several profile characteristics, these are added after each other to the left of the group name. All names at series level start with small letters. A soil classified at series level may for example be placic gleyey Typipodzol, where placic and gleyey are series designation. Theoretically, there is no limit in number of terms at series level. As yet, 20 profile characteristics have been defined, but presumably more will follow.

Definition of soils at phase level

When classifying soils at phase level, the pH measured at about 1 m depth and the prevalent texture at 0-40 and 80-120 cm depths are diagnostic terms. Unlike the 3 other levels, there is no special nomenclature of profiles at the phase level, texture and pH being stated just to the right of the profile name.

Storage of soil classes in a computer

At the end of 1983, about 9000 soils had been classified along the main pipeline in Denmark, and for long distances soils had been classified every 25 metres. All these data up to series level are stored in the computer in numerical form, and the position of the profiles are indicated by UTM coordinates. In this way, it is possible to interact the soil classifications with the analysed soil profiles and the soil maps.

Table 2 shows examples of soil types designated in numerical form.

Table 2. Different soil types described in numerical form.

	Series	*Group*	*Order*
gleyey placic Typipodzol	07,13	0,0,1	06
Humøs Kalktypigley	03	6,0,1	09
Pseudogleykalkbrunjord		0,13,6	04
kolluvial Blandingsbrunsol	01	0,0,3	03
placic Podzolstagnogley	13	0,0,26	08
pseudogleyey Typilessive	09	0,0,1	05
histic Lessivevadgley	04	0,25,2	09

The numerical classification system makes it possible to find all soils with certain characteristics. Soils having cemented layers are found by searching for number 14 at series level, while all soils having A1 horizons thicker than 40 cm are found by searching for
- number 7 at order level
- number 9 at order level in combination with number 4 at group level
- number 1 at series level

Soils with degraded Bt horizons are found by searching for
- number 5 at order level in combination with number 4 at group level
- number 11 at series level

Soil classifications in relation to planning and science

The soil classification data from along the trench will be widely used in the physical and agricultural planning. The data supplement, the soil maps, for instance on pedogenesis, variation in texture and drainage classes. The combination of many classification data made directly in the field and some data from profiles analysed in detail for water retention and root development provide a far better basis for water planning than today.

At the Bureau of Land Data, a national registration of soils with an actual or potential need for drainage is carried out. In this respect, the 8000 data sets from the trench are of great importance because they make it possible to calculate the portential need for drainage within the mapping units of the soil maps. This is done by calculating the percentage shares of classified profiles, within a mapping unit, having gley, stagnogley or pseudogley at different levels according to the classification system.

The soil classification data extend knowledge, for instance, about pedogenesis in different parent materials and the variation in soil types in relation to land-forms.

Summary

A hierarchical system of soil classification with four levels has been developed for classifying soils directly in the field. By means of this system, about 9000 soils are being classed in connection with the establishment of the main gas pipeline in Denmark.

The system is based on easily detectable diagnostic horizons and profile characteristics indicating the horizon sequence. All the classifications are stored in a computer at the Ministry of Agriculture's Bureau of Land Data. In this way, it is possible to interact the soil classifications with digitized national soil maps prepared in 1975-1980.

References

FAO/UNESCO, 1974. Soil map of the world, Vol. 1, legend. Unesco, Paris.

FAO, n.d. Guidelines for soil profile description. FAO, Rome.

Madsen, H.B., 1982. Profilbeskrivelsesvejledning. Landbrugsministeriet, Arealdatakontoret, Vejle.

Madsen, H.B., 1983. A Pedological Soil Classification System for Danish Soils. Ministry of Agriculture, Bureau of Land Data (in print).

Soil data and decision making in agriculture

A.N. Sadovski, N. Poushkarov Institute of Soil Science, 5 Shosse Bankya, BG-1080 Sofia, Bulgaria

This paper gives some examples of using digital soil data in agriculture. In order to expand agricultural production and meet demands for food it is necessary, in the long run, to increase crop production from each particular field. Such an increase can be ensured only if the following factors are fully taken into account: agro-ecological conditions, characteristics of each field and the control of the processes contributing to yield in that field, corresponding to the soil properties, biological peculiarities of the crop and variety grown and the actual conditions of production. That is why there is a need to establish systems to process vast amounts of soil, climatic and other data and to supply farmers with information suitable for management and technical decisions.

Agriculture in Bulgaria has developed rapidly over the last 35 years. The average yield of wheat has increased from 1.2 to 5 t/ha and that of maize from 2 to 6 t/ha. Capital investment in agriculture is several times higher than in 1948, with NPK fertilizer input up to 250 kg/ha. The area under irrigation has increased from 30 000 to 1 200 000 ha. The average farm size, which was 0.6 ha in 1948 has, with the creation of cooperative farms and the establishment of agro-industrial complexes (APKs), increased to 27 000 ha (270 km^2).

In Bulgaria, agriculture is practised with a great variety of soils and climatic conditions. The country comprises so many soil types and varieties that it could be called the 'Soil Museum of Europe'. Of agricultural importance are all the lands situated between 0 and 1500 m above sea level. The country's diverse landscape and broken relief determine the great variety in the climatic conditions.

The first steps in using digital computers for solving real technical problems in agriculture were made in the early 1970s when a model was developed for giving fertilizer recommendations (1st approximation) on the basis of results from large-scale soil survey and chemical analysis (Garbouchev & Sadovski, 1974; Garbouchev, 1974). The necessary fertilizers for 36 main crops were determined by this model with a computer. Then came a model for scheduling irrigation of 6 main crops (Krafti et al., 1980). The two models were in the form of program packages and made use of soil, climatic and reference data organized in separate data files.

Gradually, as the users' requirements for a complete, precise and timely information increased, the idea was conceived of an integrated computerized agriculture management system (ICAMS) based on an agro-ecological databank (Garbouchev et al., 1978; Garbouchev & Sadovski, 1978; Sadovski, 1981). The system is called 'integrated' for several reasons.

- The problems concerning optimal organization and technical management of agriculture will be solved hierarchically at three levels - long-term planning and management, annual technical and economic planning and management, and current planning (scheduling) and control.
- The system comprises the following subsystems - methods and models, algorithms and software, database, computer hardware, engineering projects and technical know-how.
- The database includes soil, climatic, biological, agro-technical, socio-economic and experimental data.

The aim of such system is to solve two problems: prediction of yield of the main crops grown in the country and making of technical decisions to obtain such yields.

The estimation of the expected obtainable yields (EOY), limited only by the actual soils and climatic conditions, is based on an assessment of land productivity. The assessment is made according to crops, for instance wheat, maize, sunflower, sugar-beet, cotton, vine, tobacco, tomato and lucerne, taking into consideration soil and climatic data. The relative grading (with values from 0 to 100) made on the basis of soil properties that have the greatest influence on crop growth, namely texture of topsoil and lower subhorizons, thickness of humus horizon, soil thickness, texture coefficient, humus content, soil reaction (pH) and watertable, is adjusted for extent of erosion, salinization and stoniness. The field grading value (FGV) is calculated by multiplication and takes into account some climatic factors such as natural water supply, sum of temperature above 5 °C, sum of temperature for growing period and precipitation during growing period.

Calculation of expected obtainable maize yield

$$EOY_i = FGV_i \cdot UTF \cdot CA \cdot CR$$

where:

EOY_i = predicted maize yield on field i
FGV_i = field grading value
UTF = unit transformation factor ('price' of 1 FGV in kg/ha)
CA = coefficient of altitude (= 1 at sea level)
CR = coefficient for reduction of yields (from experimental to farm conditions)

Similar equations have been derived for other main crops. They are based on the assumption that yield is proportional to PGV. They give a good estimate of the yield under average climatic conditions (in 80 % of the fields). For the remaining 20 %, low and high limits of the expected yield are determined on a probability basis.

Precise calculation of FGV depends to a great extent on firm knowledge of the soil properties. For the majority of cases in soil studies, the values of different soil characteristics are determined at unequal intervals along the profile depth, corresponding to different genetic horizons and only rarely are they determined at equal intervals of 10-20 cm. Besides, their values at the boundary points of the profile (the surface and the lowest point of C horizon) are often not known. These difficulties can be overcome by describing soil properties as functions of profile depth by cubic splines of the kind

$$s(x) = y_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$$

$$x_i \leq x \leq x_{i+1}, \quad i = 1, 2, \ldots, n - 1$$

where b_i, c_i, d_i are coefficients of the spline in the interval $[x_i, x_{i+1}]$. Extrapolation of the soil properties at the boundary points is performed by conversion of the inpterpolation formula of Newton for unequally spaced values of the argument and using constraints for the different soil properties (Sadovski, 1982).

The obtained splines are used for different purposes such as:
- Determining values of soil properties at any desired point in the profile,
- Determining the mean value of soil properties in a horizon of interest: topsoil (0-30 cm), subsoil (30-100 cm), active soil (0-100 cm)

$$M = \int_{\alpha}^{\beta} x \cdot s(x) \cdot dx$$

- Determining the total quantity of a given substance in a certain soil layer by integration

$$Q = \int_{\alpha}^{\beta} (x) \cdot dx$$

As illustration, Figure 1 shows the soil reaction (pH) and clay content (%) of four selected profiles.

The model for calculating fertilizer recommendations (2nd approximation) gives the amounts of the following nutrients: nitrogen, phosphorus, potassium, magnesium, lime and the trace elements boron, molybdenum, zinc, copper and iron, required for the target yield on each particular field. The N fertilizer requirements are determined on the basis of N removal by plants and then corrections are made for previous crop grown, humus content of soil, erosion and other factors. Rates of P and K are aimed at building up optimum levels of nutrients in soil (Poushkarov, 1980).

Calculation of N rate:

$$RN_{ijm} = (a_{jo} + a_{j1} \cdot Y_{im} + a_{j2} \cdot Y_{im}^2) \cdot C_{ijm}$$

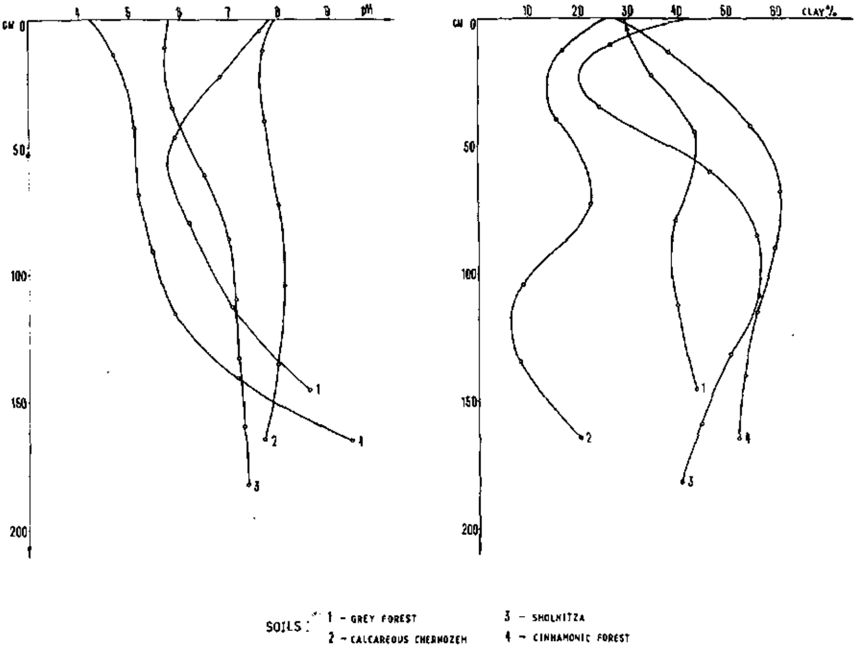


Fig.1. Soil reaction (pH) and clay content (%) of four selected profiles. 1: grey forest, 2: calcareous chernozem, 3: smolnitza, 4: cinnamonic forest.

where:

RN = rate of nitrogen
i = code number of the field
j = soil type or soil group
m = crop grown
Y_{im} = target yield
C_{ijm} = correction coefficient
a_{jo}, a_{j1}, a_{j2} = regression coefficients

Calculation of P rate:

$$SP_{ij} = A_j \cdot \exp(B_j \cdot Z_i^2)$$

$$RP_{ijm} = \left(\frac{SP_{ij}}{T} + U_{ijm}\right) \cdot k_j$$

where:

RP = rate of phosphorus
SP_{ij} = sorption capacity (amount required to reach optimal concentration of 0.2 mg/kg P in 0.01 mol/l $CaCl_2$)
A_j, B_j = regression coefficients
Z_i = available P established by soil testing
T = time to build up a phosphate level (= 12 years)

U_{ijm} = amount of P taken up by crop
k_j = coefficient of P ageing

Calculation of K rate:

$$RK_{ijm} = \frac{U_{ijm}}{K_1} + \frac{(Z_j - Z_i) \cdot K_2}{K_3 \cdot T}$$

where:
RK = rate of potassium
U_{ijm} = amount of K taken up by crop
K_1 = coefficient of K utilization
K_2 = coefficient of calculation for bulk weight of plough layer per hectare
K_3 = coefficient of the effect of newly applied K
Z_j = optimal level of K required
Z_i = available K established by soil testing
T = time to build up a potassium level (= 12 years)

The fertilizer recommendations given in the form of computer printouts assume that the estimated rates for each particular field or that fields with similar values of fertilizer rates and application of averaged rates be closely followed. Averaging is applicable when the fields are small, or aircraft and high-productive machines are used for fertilizer application. After calculating the fertilizer rates, soils are classed by nutrient requirement with a numerical classification in which elements are the fields in an agro-industrial complex (APK) and variables are rates of N, P and K. The system uses cluster analysis by Ward's method and by average linkage method, using similarities computed by Euclidean distance. The result is a dendrogram showing the classes and the fields they include.
It is worthwhile to give some data on the size of the database. The following Table contains files which it includes at present as well as number of the characteristics (attributes), number of records and size of the files in bytes.

File	Number of attributes	Number of records	Size in bytes
Soil	37	$3.5 \cdot 10^4$	$4.5 \cdot 10^6$
Climatic	22	$1.5 \cdot 10^3$	$1.1 \cdot 10^5$
Biological	20	$1.1 \cdot 10^5$	$7.5 \cdot 10^6$
Agrotechnical	25	$1.1 \cdot 10^5$	$9.5 \cdot 10^6$

The functioning of the described computerized management system is accomplished by a Data General dual NOVA-ECLIPSE mini-computer system. The establishment has begun of an interactive system linked by telephone. In that way, the production units (APKs, brigades) scattered throughout the country will be able to use central collected and processed data.

The implementation of ICAMS is made at several stages including the separate subsystems and gradual coverage of wide areas of arable land. Thus, for example, in 1975 the system served 1.6 million hectares, in 1982 3 million hectares and it is envisaged that by 1995 all farmland amounting to 5.5 million hectares will be covered by this system.

The implementation of an integrated management system with computers, mathematical models and optimization techniques is prerequisite to achieve high economic results in agriculture. In 1980, the direct economic effect of the presented system amounted to 3.5 million Bulgarian leva. (At official exchange rates, 1 US$ = 0.97 lev (1983)). At the final stage of implementation in 1995, benefit is expected to reach the considerable sum of 530 million leva. The solution of one of the most important socio-economic problems - meeting the population's food demands - depends to a great extent on timely development and use of the integrated computerized agriculture management system. This is the only way to achieve an optimal use of the country's limited land resources and to intensify agriculture.

References

Garbouchev, I, & A. Sadovski, 1974. Agric. Science, v. XIII, No. 2, p. 111, Sofia (in Bulg.).

Garbouchev, I. (Ed.), 1974. Control of Soil Fertility. p. 492. Zemizdat, Sofia (in Bulg.).

Krafti, G. & A. Sadovski et al., 1980. Plant Science, v. XVII, No. 4, p. 135, Sofia (in Bulg.).

Garbouchev, I., A. Sadovski & M. Benevski, 1978. In: Developments in Soil Information Systems (A. Sadovski & S. Bie, Eds), p. 4. Pudoc, Wageningen.

Garbouchev, I. & A. Sadovski, 1978. In: 11th Internat. Congr. Soil Science - Symp. session papers, v. 3, p. 132, Edmonton.

Sadovski, A., 1981. Agric. Science, v. XIX, No. 4, p. 34, Sofia (in Bulg.).

Sadovski, A., 1982. In: 3rd National Confer. Soil Science - Proc., v. 1, p. 76, Sofia (in Bulg.).

Poushkarov, N., 1980. Engineering project for production of expected obtainable maize grain yield in P.R. Bulgaria during 1980. Inst. Soil Sci., Sofia (in Bulg.).

Compression of soil data for regional problem analysis

L.B. Pachepskaya & Ya.A. Pachepsky, Institute of Soil Science and Photosynthesis of the Soviet Academy of Sciences, Puschino, Moscow Region, USSR

Introduction

The analysis of regional problems requires generalization of point data. Compressed data should represent peculiarities of the region as well as problem peculiarities. Proper generalization produces a model. The model allows us to predict and to study the changes occurring in the region, including those unobserved earlier.

The databases of the soil information systems give the necessary information for the generalization. However the generalization methods set certain requirements upon the structure and contents of the data. A few examples of such interaction are considered below. One of the groups of generalization methods comprises the similarity theory methods.

Case I

The region is the southern part of the Ukraine; the problem is the evolution of soil and hydrogeological conditions owing to irrigation. The database includes the preliminary study data and the soil survey data. The practical aim is to estimate upflow of groundwater into the aeration zone during the growing season with respect to watertable. Locally this task has been solved by regressions analysis. For generalization, groundwater upflow Eg, evapotranspiration Ec and watertable H have the dimension of length. Then according to dimensional theory, it is possible to introduce two dimensionless parameters sufficient for the complete description of the phenomenon (if the initial system of the parameters Eg, Ec and H includes all the essential parameters). In fact, having introduced the dimensionless parameters $\pi_1 = Eg/H$ and $\pi_2 = Ec/H$, we found that they are definitely dependent on each other and the dependence proves true throughout the whole region.
When assessing salt concentration of groundwater C during upflow, the following parameters have been adopted as essential: the amount of supplied water Q; the increment of watertable ΔH and the salt concentration of the water held in the soil γ. According to theory, there should be two dimensionless parameters. Having introduced $\pi_1 = C/\gamma$ and $\pi_2 = Q/\Delta H$, we discovered a close relationship between them.

In the same way, having adopted four dimensionless parameters we obtained relationships to be used for calculation of soil salinity increments.

Case 2

The region is the lower reaches of the Amu Darya (River Oxus); the problem is impact on soil and hydrogeological conditions of irrigation. The database includes features of relief and of soil cover including patent material, hydrogeological and hydrochemical conditions. The task is prediction of salt concentration of water draining from reclaimed areas. Dimensional analysis shows that one can introduce seven dimensionless complex relationships to determine effluent salinity. These relationships remain valid only within the certain periods of time corresponding to reclamation stages during which the nature of improvement does not change. We found a piece of land similar to the territory under study; the only difference was that reclamation had been started some time earlier. The values of the dimensionless factors at the early stages of land reclamation were approximately the same for the both territories. The idea was that the approximate coincidence of the complex values for both territories would prove true also during further stages of land reclamation. The dimensionless values were used to predict the dimensional values of interest, in particular effluent salinity.

In general, such methods provide ways for generalization. However if the values of some essential parameters are unknown, it is impossible to find the relationships between the dimensionless complexes or similarity of the territories cannot be proved. That is why we should make a joint analysis of dimensions of the parameters to be included in the database and which are to be obtained as a result of modelling.

Another group of generalization methods is based on routine models of soil features or processes, that is, on the models tested for application under various conditions.

Case 3

The region is the Hungarian Plain; the problem is hydrophysical factors of soil productivity; the database comprises physical and chemical parameters of soil as well as hydrophysical functions; the task is to use data on the soil texture and bulk density for evaluation of waterholding capacity and hydraulic conductivity. The routine model is a logarithmically normal distribution of the effective radiuses of pores. It follows from the model that the water retention curve is given by the equation $q = \frac{1}{2}$ inverf $b(\mathrm{pF} - \mathrm{pF}_*)$, where $q = (\underline{\theta} - \underline{\theta}\mathrm{res})/(\theta\mathrm{o} - \theta\mathrm{res})$, $\underline{\theta}$ is the volume feaction of water, $\underline{\theta}_\mathrm{o}$ is porosity, θ_{res} is the volume fraction of residual water, b and pF_* are empiric constants which can be used to express moments of the frequency distribution function. The routine model also shows that the unsaturated hydraulic conductivity is expressed in terms of imcomplete beta functions: $k - MB_9(-2/b+1, 2/b+1)$ accurate to the matching factor. Statistical analysis shows that in the region considered the water-holding capacity of genetically similar soils is determined by bulk density at low pF and fine clay content at high pF. At intermediate pF, the water retention curve is satisfactorily approximated by the above equation in which b and pF_* are found from the data only in the intervals of low and high pF. The water retention curve found allows us to find the hydraulic conductivity curve. Within the region considered, we mark out contours inside which the values of b, pF_* and M vary much less than beyond the boundaries of the contours.

Case 4

The region is the north and centre of the European part of the Soviet Union. The problem is decomposition rate of the pesticide (3,4-DCA); the database comprises properties of soils and soil regions as well as dynamic curves of pesticide decomposition; the task is to determine the influence of soil conditions.

The routine model is the logistic law of decomposition dynamics: $c(t) = c_0/(1+(t/t_{0,5})p)$, where c_0 is the initial content, t is the time, p and $t_{0,5}$ are empirical parameters. Estimation of average values of p and $t_{0,5}$ and their dispersions leads to the conclusion that p is independent of the soil type, and $t_{0,5}$ depends on soil conditions. The contours of constant $t_{0,5}$ are plotted.

Conclusions

Thus application of the routine models contributes greatly to generalization of local data. It leads to marking out the contours within which model parameters do not differ significantly. However the database should provide information not only to identify the routine models but also to verify them. As routine models, we can use thermodynamic equations, the equation of plant biomass dynamics under the influence of photosynthesis and respiration as well as other equations.

For generalization, we must usually choose one model from several. A single criterion of adequacy is insufficient. We use the system of criteria that allows for calculation of residual dispersion, limits of the confidence interval, the autocorrelation coefficient of residuals and the standard error of model parameters; the system also allows use of the Williams-Kloot criterion. A loss of detail during generalization is compensated for by the compactness of presentation and by the possibility of comparing the model's factors that allows classing of soils into categories according to the considered dependences.

If the model obtained by generalization is used for prediction, the prediction should necessarily be polyvariant. The predicted values should be considered as random ones. To evaluate the moments of their frequency distributions the data sufficient for computer simulation of random input values are included in the database.

Future development of the generalization methods can only be achieved with databases comprising information necessary to test and compare these methods.

Approach to soil survey and computer-aided land evaluation for land consolidation in Turkey

S. Şenol, University of Çukurova, Faculty of Agriculture, Department of Soil Science, Adana, Turkey
P.A. Burrough, Agricultural University, Department of Soil Science and Geology, Wageningen, the Netherlands

Introduction

In this paper, we aim to show how a proposed process of land evaluation for land consolidation and reallotment in rural areas of Turkey could be made easier and more efficient with computer-aided methods. We believe not only that computer methods could speed the processes of interpretation and evaluation but also that because they are reproducible and definable, their use would also make the results of reallotment more acceptable to the farmer.

Because of the land ownership laws of Turkey, land holdings have been divided into many small parcels as inheritance passed from one generation to another. Some 52.2 % of all land parcels are smaller than 3.1 ha and 96.3 % are smaller than 20.1 ha. Moreover, many parcels are irregular in shape and not suited to mechanized agriculture (Çalgüner, 1970). The first land consolidation project started in 1961 for the village of Karkin in Konya; later attempts were made in Antalya and the surrounding areas (Ercan, 1966). These early attempts at land consolidation were based on land indices calculated according to the Storie Index (Storie, 1948). Because this assessment placed too much weight on purely physical aspects of the land, its use in assessing the economic value of land parcels was strongly disliked by farmers, resulting in legal proceedings and a discontinuation of land consolidation activities. Clearly, a more integrated approach to assessing land values is required, but this increases enormously the complexity of the land consolidation task. Here the computer can be of tremendous help, though we shall in no way claim that it will ever be able to dispel the farmer's healthy disrespect for governmental agencies!

System of land evaluation

The main aim of this study is to bring a new approach to land evaluation that allows computation of a land index for each parcel of land as the basis for reorganizing them in the land consolidation projects. The land evaluation system proposed is based mainly on the principles given by 'A Framework for Land Evaluation' (FAO, 1976). It is a quantitative and specific-purpose land-evaluation system at the detailed level of intensity and refers to the current suitability. Because of the large amount and complexity of data required, each mapping unit and land parcel can only be separately evaluated if the land evaluation system is computer-assisted. Information collected during the soil survey, agronomic and economic

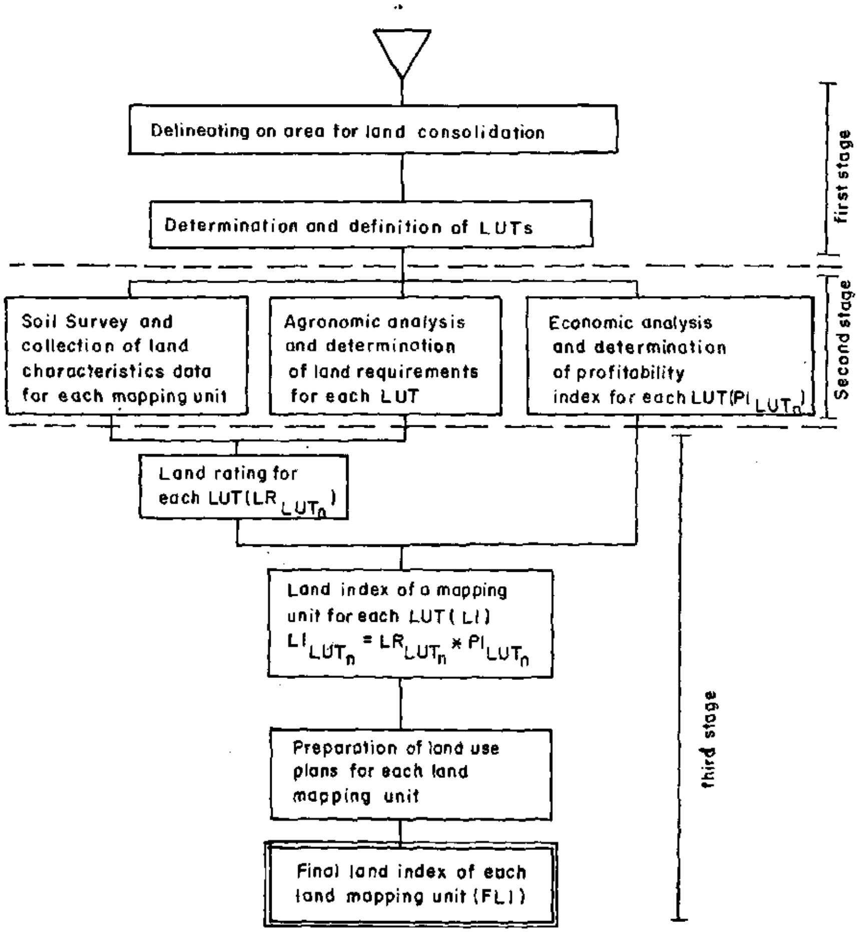


Fig. 1. Flow diagram of the whole land evaluation procedures. *LUT*, land utilization type; *LR*, land rating; *LI*, land index; *PI*, profitability index; *FLI*, final land index.

analysis are stored in the computer and processed for the calculation of the final land index. The final land index is a site-specific value and a comparative value of the land with respect to the others.

Stages of the proposed land-evaluation system

The procedures of the proposed land evaluation system are divided into three stages (Fig. 1):
1. feasibility studies,
2. data collection,
3. evaluation.

Stage 1. Feasibility studies

Delineating one area for land consolidation. The areas chosen for land consolidation must be considered as a unit. The exchange of land can only occur within this unit. Land is evaluated only in terms of the differences found within the area in which parcels of land are subject to exchange. This area is unlikely to embrace more than one climatic zone. In order to simplify the land evaluation procedures, and because of the absence of sufficient climatic data for the rural areas of Turkey, the influence of climate will be ignored. Hence, this land evaluation system cannot be directly applied to areas where different climatic zones occur.

Determination and definition of land utilization types (LUTs). Land evaluation involves relating mapping units to land utilization types (LUTs). LUTs should be relevant to the general physical, economic social conditions prevailing in the area (FAO, 1976). Previous land-use and farming system investigations of area must be taken into account when identifying relevant LUTs (Beek, 1978). For that reason, the present LUTs in the area and its surrounding area within the same climatic zone are determined. The LUTs are defined at detailed level and in terms of the following key attributes: produce, labour, capital, farm machinery, management, scale of operation (Beek, 1978). These key attributes are described in quantitative terms.

Stage 2. Data collection

Soil survey and collection of data on land characteristics. A special-purpose soil survey will be carried out, but it will contain at least as much general information about soils as general-purpose soil surveys. The soil survey needs to be large scale (e.g. 1 : 10 000) and requires a high density of observations and narrowly defined soil classes. Topographic maps and aerial photographs will be used as the basic cartographic materials to identify land mapping units.

The proposed soil survey would be done in three stages (Fig. 2). In the first stage, soil surveyors would make auger-hole observations at regular spacings. The required data would be recorded on standard forms which would be designed to capture accurate and objective data quickly. Standard data collection and description methods are also essential for computer manipulation of soil data.

The second stage of the soil survey is the processing of the information gathered from the auger observations in order to define map units and legend classes. The soil properties can be grouped into three kinds: nominal, ordinal (or ranked) and metric (or continuous) (de Gruijter, 1977). The histograms, ranges, means and standard deviations, and other appropriate summary statistics can be used to help determine the most suitable class devisions and model concepts. The calculation of basic statistics can most easily be done by computer (Rudeforth, 1975), but more advanced techniques

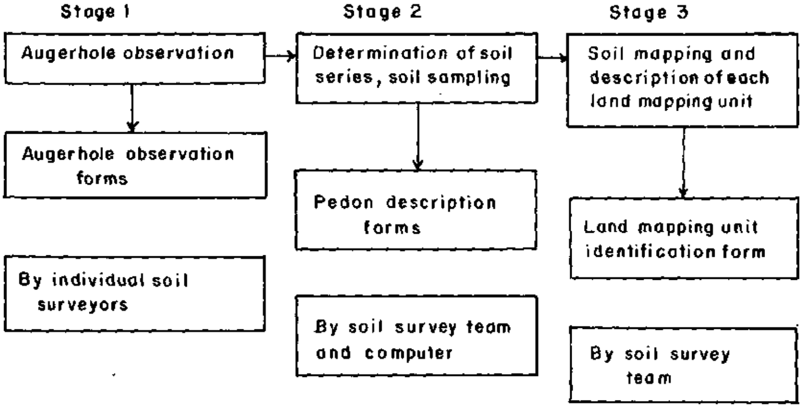


Fig. 2. Stages of soil survey.

of numerical classification and optimal allocation could be used to achieve the best classification as quickly as possible (e.g. Burrough & Webster, 1976). The classification rules can then be written out as a key, suitable for use in the field or in the computer.

The third stage involves classifying each auger-hole observation and delineating on the aerial photo or map the resulting mapping units. A mapping-unit identification form is filled in for each unit. This would be held in the computer in a file of basic map-unit data so that the information collected during soil survey can be used as basic information for rating the indices of land mapping units and land tracts.

Agronomic analysis and determination of land requirements. For each LUT defined during the first stage, land requirements are de-

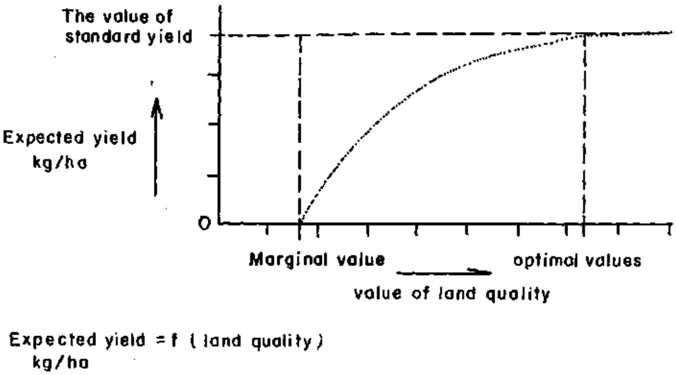


Fig. 3. Illustration of the relation between the expected yield and land quality.

termined. Optimal and marginal values of each land quality that affect the productivity of land under defined LUT are determined. The most appropriate graph showing the relation between the expected yield and land quality or characteristics are drawn and mathematically expressed (Fig. 3). The standard yield of a LUT is the maximum yield that can be obtained under defined key attributes of LUT when all land qualities and characteristics are optimal. The expected yield is the amount of produce that can be obtained when a land quality or characteristic is limiting while the others are optimum. The expected yield should increase as the limitattions decrease, but it cannot exceed a standard yield.

Optimal and marginal values of land qualities determined for individual LUTs should be as accurate as possible. Ideally, the expected yield would be obtained by evaluating the form of the regression between yields and land qualities, but in practice, more pragmatic methods will probably be needed.

Economic analysis and determination of profitability index (PI). The necessary financial inputs (Turkish lira per hectare) for all described procedures of each LUT are calculated according to the prices of these procedures. The sum of all expenses made in order to obtain the standard yield under optimum land conditions is the total input. The output is calculated by multiplying standard yield by market price (Turkish lira per kilogram) of produce. The benefit *(B)* obtained from LUT_j is

$$B_{LUTj} = (\Sigma \text{ output}) - (\Sigma \text{ input}) \tag{1}$$

and the profitability index (PI) for LUT_j is

$$PI_{LUTj} = \frac{B_{LUTj}}{B_{max}} \tag{2}$$

where B_{max} is the maximum benefit that can be obtained from one of the LUTs within all LUTs considered.

Stage 3. Evaluation

Rating of land qualities and characteristics with respect to land requirements of each LUT for each mapping unit. The limiting nature of each land characteristic is taken into account by its effect in reducing productivity. This limiting effect either causes a decrease in the obtainable yield or increases the necessary inputs. The expected yield (kg/ha) for the first kind of land characteristics, such as shallower soil, are obtained from the graph obtained during agronomic analysis. Other land characteristics cause extra input, e.g. position of land; these are evaluated by subtracting from the standard yield the yield needed to be sold to meet the extra inputs. So, the expected yield in the second case is

$$\begin{matrix}\text{expected}\\\text{yield}\end{matrix} = \begin{matrix}\text{standard}\\\text{yield}\end{matrix} - \left(\frac{\text{extra input}}{\text{market price of produce}}\right) \tag{3}$$

The productivity rating of index of each ecological land quality is rated according to the methods proposed in the Soil Survey Manual (Soil Survey Staff, 1951) which is

$$\text{productivity rating index} = \frac{\text{expected yield}}{\text{standard yield}} * \tag{4}$$

Because we are dealing with an existing agricultural system, but one in which the relations between the individual ecological land qualities and characteristics and crop yield have not been fully ascertained, it would be simpler, and more realistic, to rate each land mapping unit according to the super land quality 'yields obtained', rather than to any ad hoc weighted sum or product of expected yields based on the individual land qualities. Both 'standard' and 'obtained' yields could be obtained, in principle, from a number of benchmark site-yield studies over several years (McRae & Burnham, 1981).

The land rating of each mapping unit for each LUT would then be

$$LR_{i\mathrm{LUT}j} = \frac{OY_{i\mathrm{LUT}j}}{SY_{\mathrm{LUT}j}} \tag{5}$$

where:

$LR_{i\mathrm{LUT}j}$ = the land rating of mapping unit i for the jth LUT
$OY_{i\mathrm{LUT}j}$ = the observed yield of LUT_j in mapping unit i
$SY_{\mathrm{LUT}j}$ = the standard yield of LUT_j

Computation of the land indices for each land mapping unit with respect to LUTs. The $LR_{\mathrm{LUT}j}$ reflects only the physical characteristics of land. They are not comparable with other land ratings of other LUTs. In order to make land ratings comparable and to establish the most profitable land-use plan for that mapping unit, the calculated land rating of that land mapping unit for a LUT is multiplied by the profitability index of that LUT. The result is the land index ($LI_{\mathrm{LUT}j}$) of the mapping unit for LUT_j.

$$LI_{\mathrm{LUT}j} = LR_{\mathrm{LUT}j} * PI_{\mathrm{LUT}j} \tag{6}$$

Preparation of land-use plan for each land mapping unit. A farmer should not attempt to apply all the LUTs which have been considered for the land evaluation. Certainly, he should choose the most suitable and profitable LUTs for his land. From this point of view, a model land-use plan is prepared for each land mapping unit for a finite period of time. The aim of this model land-use plan is to indicate how the land mapping unit can be used to get the maximum benefit under sustained agriculture. The land-use plan is prepared for the chosen time period by taking into account the following factors:

- possible crop rotations for sustained agriculture
- growing periods of suitable LUTs
- land index (LI_{LUT}) value of suitable LUTs.

Computation of the final land index (FLI). The final land index is the value of a land mapping unit relative to other land mapping units. The model land-use plan prepared for a land mapping unit will have a basis for the calculation of the final land index. The land index of each LUT that has been used in the land-use plan are added and the sum is divided by the number of growing periods during the finite time period of the land-use plan (each occurrence of a LUT is assumed to operate for one growing period).

$$FLI = \frac{\sum_{i=1}^{N} \sum_{j=1}^{P} LI_{\mathrm{LUT}ij}}{M} \tag{7}$$

where:
N = number of LUTs in the land-use plan
P = occurrence of a LUT in the land-use plan
M = total number of growing periods in the land-use plan

Index rating for tracts of land

The index rating for a given parcel or tract of land can be obtained by weighting each final land index according to the area of the parcels covered by each mapping unit. The exact details of this procedure have yet to be worked out.
The index rating can be most easily calculated by digitizing the boundaries of tracts and the boundaries of the mapping units occurring. Before reallotment, the parcels will be small and will most likely only be covered by one or two mapping units. The procedure can be repeated to show the index rating expected after land reallotment. In this way, the farmers can see what they are likely to gain through the proposed reallotment.

Discussion

Many parts of the proposed system of land evaluation are impossible without a computer because of the large number or the complexity of the calculations. Until recently, the cost of the computer and allied equipment and staff would have been prohibitive. This would still be true if it were planned to hold all the necessary data in the computer and to perform all the intersectioning of soil and land-use maps automatically. Much of the cost of a soil information system is associated with expensive systems for disk storage, database management and with automated cartography. In these proposals, large databases would be avoided by using statistical analysis and identification keys based upon the results of such analysis to reduce to a minimum the data on soil needed for the land evaluation calculations. Polygon overlay problems (one of the best known problems in automated cartography) could be avoided by interactively digitizing the boundaries of land parcels and soil boundaries on an individual basis at the time land indices are to be calculated.

We estimate that most of the computing required for the procedures described here could be handled by a powerful 16-bit microcomputer with 256 k random-access memory, disk-storage of some 10-20 Megabytes, with two terminals, printer, plotter and a digitizer. This apparatus would have to be supplemented by a tape drive for off-line back-up of programs and data. The total cost could be in the region of US $25 000 to US $35 000, a sum that is not so very different from that required for a tractor or other large agricultural machinery. Although much software would have to be specially written for the land reallotment, these calculations are not complex, and are applicable to land evaluation in other lands. Many standard programs for numerical classification and basic statistics exist, and their incorporation into cheap, readily available packets for development work is greatly to be welcomed.

The greatest problems in implementing the system stem from the current lack of experience in quantitative land evaluation. The use of present land-use types makes it easier to define and determine land requirements than if all possible LUTs were considered. In spite of that, there will be difficulties in gathering all the necessary data. Changes in prices are assumed to act similarly for everything in the market. Some abnormal changes in the prices of goods may cause errors when estimating the profitability indices for the land utilization types. Those errors will be minimized by taking the average value of the profitability index for the last 5 years. Other difficulties may arise when rating land qualities and characteristics because some land qualities are undoubtedly the result of non-linear interactions of several land characteristics. Although these are difficult problems, it should be possible to find more acceptable solutions once they can be handled quantitatively. It should be our aim to show that these solutions are also more acceptable to the farmers and will encourage them to participate in the much-needed land reform so essential for improving agricultural production in Turkey.

References

Beek, K.J., 1978. Land evaluation for agricultural development. Publ. 23, Int. Inst. for Land Recl. and Improvement, Wageningen.

Burrough, P.A. & R. Webster, 1976. Improving a reconnaissance soil classification by multivariate methods. J. Soil Science 27: 554-571.

Çalgüner, Ċ., 1970. Arazi Mülhiyet Rejimi ve Türkiye'deki Durum A.Ü. Ziraat Fakültesi Yayinlari 437, Bilimsel Arastirma ve incelemeler: 265. Ankare Universitesi Basimevi.

Ercan, F., 1966. Türkiyese Arazi Toplulastirmasi (Şimdiye kadar yapilan calismalar ve kasilasilan problemler). Doguş Mat. ve Tic. Lim. Şik matbaasi, Ankara.

De Gruijter, J.J., 1977. Numerical classification of soils and its application in soil survey. Pudoc. Wageningen.

FAO, 1976. A framework for land evaluation. Soils Bulletin N32: FAO, Rome and Int. Inst. Land Recl. Impr. Wageningen, Publ. 22.

McRae, A.G. & C.P. Burnham, 1981. Evaluation. Oxford University Press, 239 pp.

Rudeforth, C.C., 1975. Storing and processing data for soil and land use. Soil Science 26: 155-167.

Soil Survey Staff, 1951. Soil Survey Manual, Washington United States Department of Agriculture Agric. Handbook 18.
Storie, R.E., 1948. Revision of the soil rating chart. University of California Agricultural Experimental Station, Berkley, California, USA.

Construction of a method base system for evaluating and planning of land and soil resources

J. Hatje & J. Lamp, Institute of Plant Nutrition and Soi Science, Christian-Albrecht-University, Olshausenstr. 40, D-2300 Kiel, Federal Republic of Germany

Common problems in handling soil data and programming

A working group of the institute has had various research projects on numerical methods for soil survey, classification and regionalization for more than ten years (summarized by Schroeder et al., 1980; Lamp, 1981). Therefore several local and federal collections of soil data have been assembled from trial areas and archives. These data were processed with commercial or self-made programs written and tested for special soil evaluation purposes.

The usual methods of handling soil data and programming raise problems as follows:
- heterogeneity of data in formating and coding, and of programs, for instance for internal core addressing or external job controlling;
- difficulties in linkage of data sets and of program modules for changing uses;
- poor documentation limits updating and use of old data and programs, which are remote from the original surveyor or programmer.

Emphasis in this article is given to the construction of a method base system (MBS), which aims to back a homogeneous set-up and updating of programs (methods) and a professional documentation as well as a flexible combination and use of program modules for various purposes.

Software and hardware

A main part of the KIBIS total soil information system ('Kieler Boden-Informations-System') is the DASP database system (Dorn & Kuehne, 1982), which has been amalgamated from two previous retrieval systems of the Geological Survey of the United States (GRASP) and of Lower Saxony (DASCH) (Fig. 1). As part of a Fereral research project on processing digital maps for earth resources (Vinken & Oelkers, 1980), DASP has been implemented on the Siemens-7.760 computer of Kiel University. DASP handles point data of a West German pedon data bank (Lamp, 1981) and local data collections of soil profiles and borings, including archived data of systematic land evaluations from the 1930s.

Around DASP, special programs for pedogenetic and ecological soil evaluations and for data transfer, input and output are being developed and used. These programs and commercial statistic packages

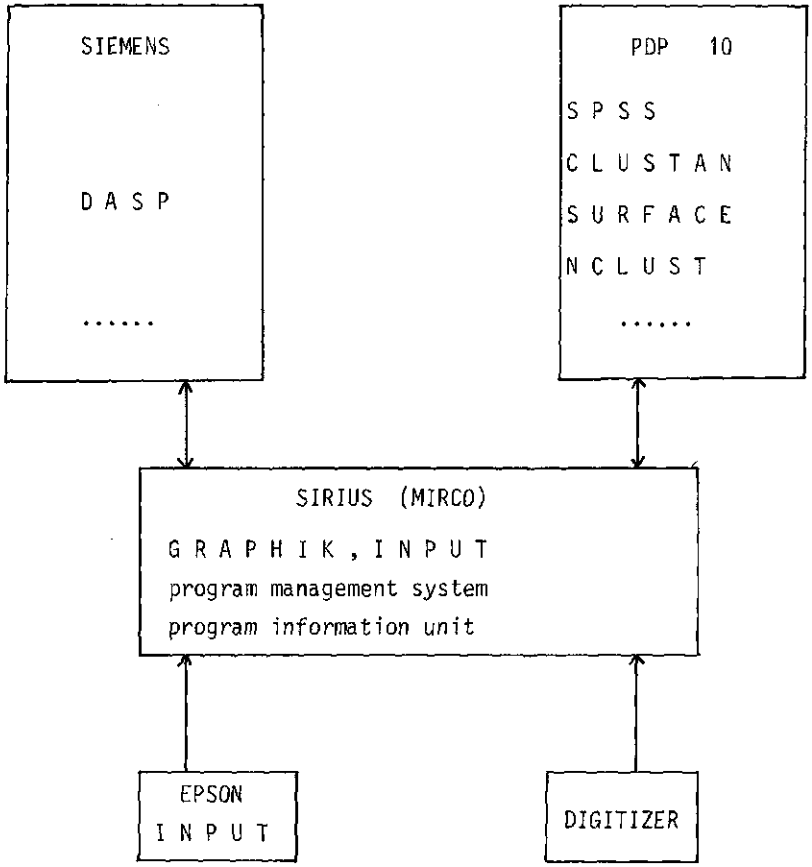


Fig. 1. Program and computer facilities of the Kiel soil information system (KIBIS).

such as SPSS and CLUSTAN, run mainly on a Dec-10 computer.

As a comprehensive program collection for spatial data analysis, especially for evaluating and planning of land and soil resources on a grid basis, TEMAKART (Volden, 1981) is being implemented as small version on a graphical 16-bit computer (SIRIUS 1) connected to a digitizing table.

Concepts, structure and tools of the method base system

Unstructured ad hoc programming usually does not allow a flexible combination of program modules for meeting changing demands of new analysis jobs. In order to avoid endless reprogramming, a base system is needed, by which subroutines, programs and program packages - here called 'methods' - may be combined and used flexibly.

Tasks of a method base system are thus:
- to set up and maintain a method library;
- to select, combine and activate program modules for given jobs;
- to guide the tasks and the execution of jobs by an interactive information system, which itself is updated by the computer.

The maintenance and running of this method base system must be comfortable and user-friendly.
To meet these aims, the method base system is being constructed with the following three parts (Fig. 1).
- A program information unit. Besides general operating instructions, information about the total system, method classes and single modules can be called up interactively at each level. The system is updated interactively.
- A program library unit allows the build-up and maintenance of a program library (add, update and delete modules) and includes a 'Command Specification and Interpretation Language' (CSAIL), by which commands for activating and structuring modules are defined and executed.
- A program management system consists of job control and security. It reads command sequences either interactively or for batch runs from a file. It allows the use of loops (REPEAT) and conditions (IF) and combines command sets to MAKROs.

Construction and use of an illustrative system

Set-up of methods and commands

The following programs may be given within a library.
DASP : database system (calling command DASP)
ZPROFI : (command ZPROFI) module for plotting soil profile depth functions by the subcommands TABINI (initializing), OPEN (file opening), ZMESSL (plots a depth scale), ZHORNA (plots horizon signatures) and ZFELD (plots fields with depth functions)
TZPROFI: module for transforming DASP into ZPROFI files (command TZPROF)
PLOTS (ND, IT, XL, YL): initializes a graphic output unit with parameter ND (logical device unit), IT (= 4, Terminal; = 9, Benson plotter), XL and YL (breadth and length of plot (command PLTINI <unit> , <length> , <breadth> , <device>)
PLOT (X, Y, IPEN): controls movement of plot pen to point X, Y
IPEN = 2, pen lifted; command GOTOP X, Y
" = 3, pen plotting; command PLOT X, Y
" = 2, pen lifted and origin shifted NEWPOS X, Y

These modules may be defined and structured to a bank by the following CSAIL commands.
INTBEG: beginning of a program bank
MODBEG: beginning of a module
CMDBEG: beginning of a command, by which modules are opened
INTEND, MODEND, CMDEND: respective closing commands
PARDEF: defines parameters for a command (names, types, ranges and standards of values)

KIBIS

System: KIBIS (name of the program bank)

Functions (commands

1. DASP

    Module 1: DASP (all database programs)

    Functions: DASP commands

2. AUSWER

    Module 2: AUSWERTUNG (all special evaluation programs)

    Functions:

    2.1 GRAPH

        Module 2.1: GRAPH (all plot routines)

        Functions

        2.1.1 PLTINI

        2.1.2 PLOT

        2.1.3 NEWPOS

        2.1.4 GOTOP

        2.1.5 ZPROFI

            Module 2.1.5: ZPROFI (plots depth functions)

            Functions: ZPROFI commands

        2.1.6 ...

    2.2 ...

3. TRANS

    Module 3: TRANSFORMATION (data transfer and transformation)

    Functions:

    3.1 TZPROFI

    3.2 ...

Fig. 2. Structure of the KIBIS program bank.

Further elements of the CSAIL language specify separation symbols (TRENN) and sequence conditions of commands (NEXT). Code opens and CODEND closes FORTRAN statements. All variables used within CSAIL must be declared (INT, REAL, CHAR).

Definition and structuring

The definition of program modules and their structuring to a method bank by using CSAIL are illustrated by the command list (Table 1). The hierarchic structure of the commands correspond to those of the KIBIS method bank (fig. 2). All retrieval functions on modules are performed as shown for the module 'ANSWER', 'GRAPH', 'PLTINI' and 'NEWPOS', respectively. Control of command sequencing is not shown, but is possible.

After definition of the bank, FORTRAN programs will be generated automatically. The standard commands (e.g. REPEAT, MAKRO) are analysed by a special interpreter and the program information unit has to be updated. The latter function has not yet been implemented. Not all problems concerning incorrect concurrencies of COMMON BLOCK names or deficiencies in core space have been solved satisfactory (e.g. Enderle et al., 1982; Fig. 2).

Examples of job execution

The powers of KIBIS in executing jobs flexibly are demonstrated by a small soil-mapping job, which includes steps of
- selecting data for a limited area by geographic coordinates;
- plotting profile depth functions of chemical and physical properties;
- plotting a map of horizon signatures.

The commands to execute this job are listed in Table 2. Parameters of commands may easily be changed to meet similar tasks. The whole list may be stored and executed repeatingly by a MAKRO command.

Results

The resulting plots of the illustration job are given in Figures 3 (depth functions of laboratory tests) and 4 (map of horizon signatures).

Summary

The construction of a method base system (MBS), including a database system (DBS), overcomes many difficulties inherent in data handling and programming: a collection of homogeneous well-structured and documented program modules (methods) are assembled in a bank and used by the system. The MBS allows convenient definition of a program bank as well as flexible analysis of data, especially for evaluation of land and soil resources and for management.

Capacities of the KIBIS method base system are demonstrated by a

Table 1. Command list for defining the program bank KIBIS.

Level	Command	Definition of
1: INTBEG	KIBIS	KIBIS program bank
CMDBEG	DASP	retrieval function on DASP database
CODE		
CALL	DASP	FORTRAN program
CODEND		
CMDEND		end of function
CMDBEG	ANSWER	retrieval function on AUSWERTUNG module
...		(as above)
CMDBEG	TRANS	retrieval function on TRANSFORMATION module
...		(as above)
INTEND		end of the KIBIS bak
2: MODBEG	ANSWER	AUSWERTUNG module
CMDBEG	GRAPH	grasp function on GRAPH module
CODE		
CALL	GRAPH	FORTRAN program
CODEND		
CMDEND		end of the function
MODEND		end of the AUSWERTUNG module
3: MODBEG	GRAPH	GRAPHIK module
REAL	XUR, YUR	two local variables
CMDBEG	PLTINI	PLTINI retrieval function
PARDEF,	DEVICE, CHAR, ('T', 'B'), 'T'	first parameter: DEVICE; type: REAL; range: 'T' and 'B'; standard: 'T'
PARDEF	LENGTH, REAL, (0:200), 30	second parameter: LEMGTH; type: REAL; range: 0 to 200; standard: 30
PARDEF	BREADTH, REAL, (0:60), 30	third parameter: BREADTH; type: REAL; range: 0 to 60; standard: 30
CODE		
XUR = 0.0		
YUR = 0.0		
IF (GERAET.FQ.'B')		FORTRAN program, which defines local variables (plot origin: XUR, YUR) and is executed only if paramaters (DEVICE, BREADTH, LENGTH) are set correctly
THEN CALL PLOTS (10,9, BREADTH, LENGTH)		
ELSE CALL PLOTS (10,4, BREADTH, LENGTH)		
ENDIF		
CODEND		
CMDEND		end of retrieval function
CMDBEG	PLOT	PLOT retrieval function
...		(as below)
CMDBEG	NEWPOS	NEWPOS retrieval function
PARDEF	X, REAL	two parameter X and Y
PARDEF	Y, REAL	
CODE		
XUR = X - XUR		
YUR = Y - YUR		FORTRAN program, which computes a new origin and moves the pen to that point
CALL PLOT (XUR, YUR, -2)		
CODEND		
CMDEND		end of retrieval function
...		
MODEND		end of GRAPH module
INTEND		end of CSAIL program

Table 2. List of commands to execute a job.

command	comment
EXEC KIBIS	Start of KIBIS program bank
DASP EINLAGERN .. SUCHEN ... ENDE	Call up DASP database and perform functions for data input, recording, defining, and executing a search for coordinate values, screening and output of the selected profile data
TRANS TZPROF ... END	Call up transformation module for transforming DASP into ZPROFI data and preparing coordinate values
GRAPH	Call up the graphic module
PLTINI, 'B', 100	Activation of Benson plotter
ZPROFI	Call up ZPROFI module
OPEN	Open data file
TABINI	Define plot sizes
ZTIEFE ... END	Call up module to plot depth values and scales, horizon signatures, depth functions etc. (Fig. 3)
NEWPOS 35, 3	Position pen to new origin (35, 3)
KRTRND...	Plot map boundaries with scales
INT X, Y BOOL EOF	Declare integer and Boolean variables
REPEAT	Beginning of a loop
READKO X, Y, EOF	Read X, Y coordinates until EOF = 1
NEWPOS X, Y	Position pen to new origin (X, Y)
ZPROFI OPEN TABINI ZTIEFE ZHORNA END	Call up ZPROFI module, open data file, define plot sizes, plot depth values (Fig. 4), plot horizon signatures
UNTIL (EOF)	Standard closing of loop
ENDPLOT	End of plot
END	End of GRAPH module
END	End of KIBIS

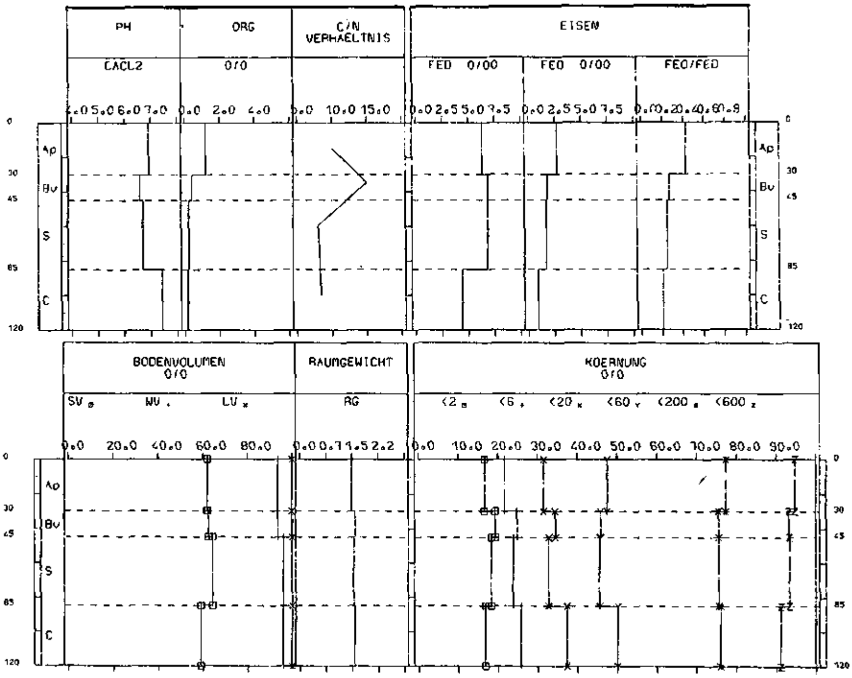


Fig. 3. Depth functions of some analytical data.

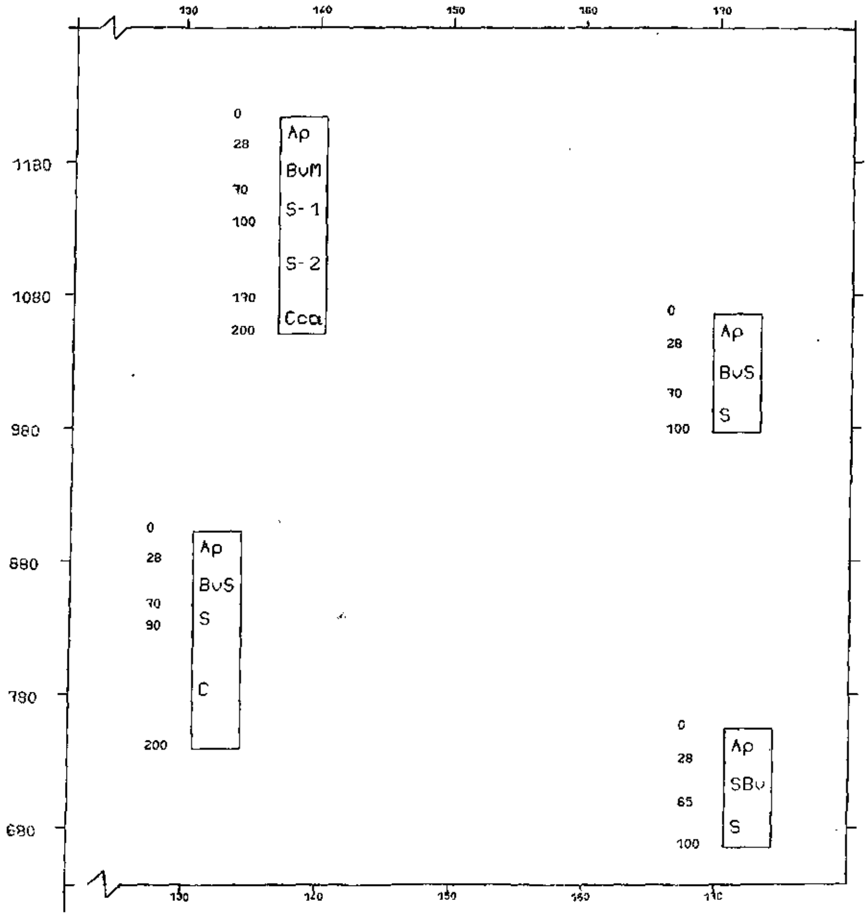


Fig. 4. Map of horizon signatures from a selected area.

small job, which plots profile depth functions and a map of horizon signatures from a small area only. At present and in the near future, programs for spatial analyses are being added from the TEMAKART program collection (Volden, 1981). This should increase the powers of both systems tremendously.

References

Dorn, M. & U. Kühne, 1982. DASP - Benutzerhandbuch. Dokumentations- und Abfrageprogramm für Schichten- und Probendaten. Niedersächsisches Landesamt für Bodenforschung, Hannover.

Enderle, G., E. Olbrich, E.G. Schlechtendahl & U. Schumann, 1982. Techniken der Modulverwaltung für Methodenbanken und CAD-Systeme. Informatik Fachberichte 57: 252-271.

Lamp, J., 1981. Morphometric analysis of pedon data. Department des sols, No. 5: 87-94. Institut National Agronomic, Paris-Grignon.
Lamp, J., 1981. Informationssysteme in der Bodenkunde - internationale und nationale Entwicklungen. Informationsverarbeitung Agrarwissenschaft 1: 113-120. Ulmer Verlag.
Schroeder, D., J. Lamp, E. Mutert, W. Kneib & J. Jakob, 1980. Soil surveying - classification - regionalisation: research projects on numerical methods. Poster, Norwegian Computing Center, Oslo.
Vinken, R & K.-H. Oelkers, 1980. Möglichkeiten des ADV-Einsatzes in der bodenkundlichen Landesaufnahme. Geol. Jahrbuch 8: 23-37.
Volden, R., 1981. Temakart, documentation and user's guide. Report No. 690, Norwegian Computer Center, Oslo.

Soil Information Systems: the problem of developing a capacity of reappraising data for single purposes

P.H.T. Beckett, Soil Science Laboratory, University of Oxford
Parks Rd., Oxford OX1 3PF, England

The problem

Soil information systems (SIS) take many forms. Essentially they are tools designed to provide the data necessary to answer questions about soil resources or soil potentialities of their area. The questions that may be asked fall into four groups.

1a. What are the values of soil or site properties A, B, C, etc. at point X (or points C_1-X_n)? This includes questions about the proportions of sites with different values of specified properties.

1b. What can I do at site X (etc.); or can I grow P or build Q, etc.; or what must I do at site X (etc.) in order to grow P or build Q, etc.?

2a. Where are there soils with the values V_A, V_B, V_C, of properties A, B, C, etc.? This includes questions about the size and distribution of areas of soil with these values.

2b. Where are there soils on which it will be possible to grow P, build Q, etc.?

We require two kinds of information to answer questions of the form 1b or 2b.

- Information about the values of soil or site properties at sites of interest.
- An 'interpretive link', that is a model or a regression that links these values to the activity under consideration, for example a model that relates wheat yield to clay content, available nitrogen, soil depth and topographic position, or a regression of bearing capacity of a soil on its content of sand, clay and iron oxide, and the quotient of cation-exchange capacity to clay content. The link may produce some form of suitability rating.

Interpretative links such as these are crucial to the use of data from SIS. Because they are often difficult to ascertain and may differ between environments or social milieux, they are often subjective. Since the success of SIS should be judged by the truth of the information it provides, and not by the fallible links through which the data may have to be used, the role of the data and the links should be clearly distinguished. The necessary set of interpretive links should be regarded as an appendage.

This being so, the information required from SIS to answer Questions 1b is the same as for Questions 1a. Questions 2b require the same information from the SIS as Questions 2a. Furthermore, to answer Questions 1a requires information for any sites that may be specified in the area covered, while Questions 2a require information *from all sites in the area in order to ascertain which of them* meet the specification in the question. The two requirements are the same, so: *SIS should be able to provide information about soil or site properties at any, or all, sites in its area.*

In practice, SIS will never be offered, nor can it hold, information from more than a finite number of sites. The sampling function will always be small - 50 pedon descriptions per square kilometre represent a sampling function of $1/10^4$, so SIS must comprise two components.

- Point data, *sets of values of soil or site properties from points,* from the small areas occupied by pedons, or the small areas from which soil samples were bulked before analysis.
- Means to interpolate (predict) values at sites between the sampled points. Since the number of unknown sites far exceeds the number of sampled sites (e.g. 10^4:1), the success of SIS largely depends upon its success in making such predictions.

Pre-computer soil information systems

Without a computer it has often been difficult to handle the volume of data produced, even by a single soil survey on medium scale. For example, in England and Wales, a soil map on scale 1 : 25 000 covers 100 km^2 and is based upon 30-60 observations per square kilometre, each of perhaps 5 site properties and 50 soil properties, which constitute some 2×10^5 values in all (Fig. 1a). Conventionally, therefore, these data have had to be generalized to make them manageable.

Classification

Usually the first stage of generalization is to produce a soil classification such that
(a) every soil in the area falls into a class
(b) every soil in a class has undergone the same genesis - if so all the soils in the class are likely to be linked to the same landscape features, and to occur in consistent spatial patterns with other classes, both of which make them easier to map
(c) every soil in a class shows similar land-use potentialities or presents similar management problems - this makes it easier to interpret the data (see below).
Sometimes Requirements b and c are not compatible, or they cannot be achieved.

Soil classes will be defined by the values of some of their properties (definitive features). It is assumed that members of a class will show some similarity in other properties too (associated features). Thus the properties of all the members of a class are summarized in the class description, that also includes its definition. Conversely the properties of one soil may be summarized by recording its class name (Fig. 1b). In the example, the information now represents 3.5×10^3 observations, together with class descrip-

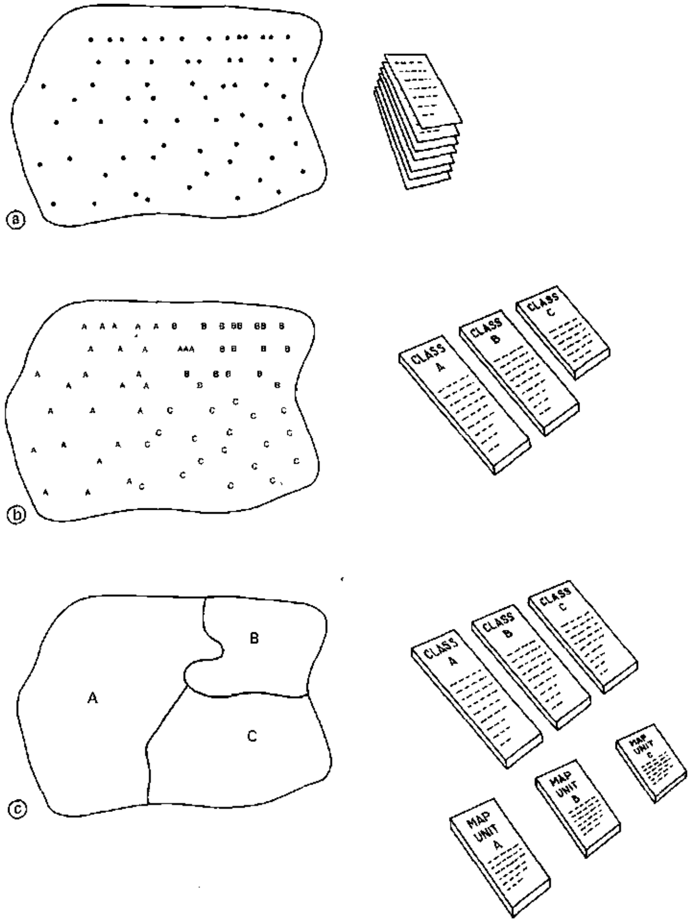


Fig. 1. Soil maps: alternative models of information storage.
(a) Each observation site may produce values of 50-60 soil and site properties, perhaps 2×10^5 values for the area covered by one soil map at 1 : 25 000 in England or Wales.
(b) If the soil at each observation site is classified, then the site properties are subsumed into a class name for each observed soil and a detailed description of each class.
(c) An area dominated by one soil class (or a repeated pattern of soil classes) may be subsumed into one map unit, whereby all the soils of the area described in a detailed description of each map unit and a detailed description of the classes in each map unit. This is an ideal - reality is not so simple.

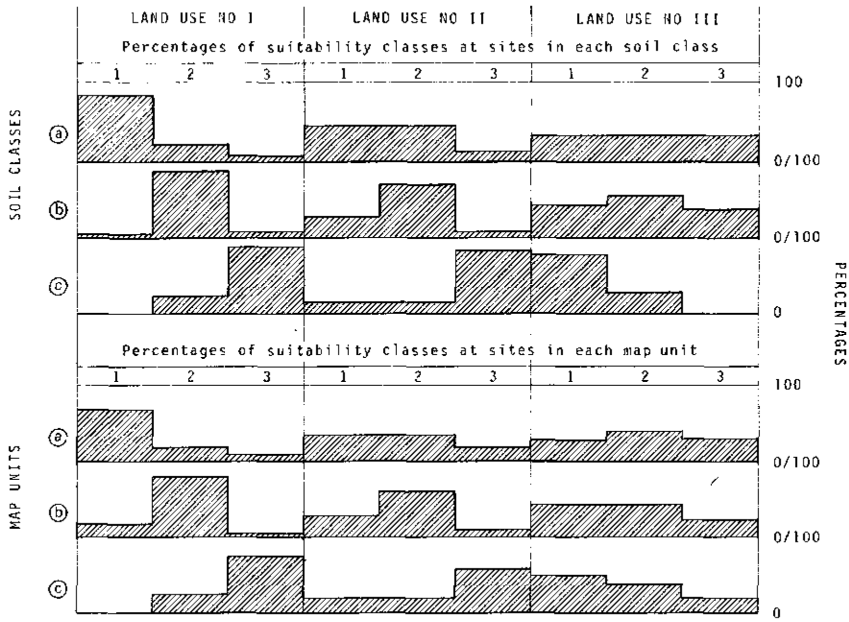


Fig. 2. The percentage of sites occupied by soil classes A, B or C, or in map units A, B and C, that fall into suitability classes 1, 2 or 3 for land uses, I, II or III. Note that the map or classification are good predictions for land use I and poor predictions for land use III. Map units usually contain some impurities, so map units may be worse predictors than soil classes.

tions.

Mapping

Next, any part of the area that is dominated by soils of the same class may be defined as a *map unit*, and distinguished (*delineated*) by a soil boundary. The description (*legend*) of the map unit describes the soils within it. Ideally map units are 'pure': 100 % of randomly sited check points will find its dominant soil class. In practice, the purity of a map unit does not usually exceed 65-75 %, and may be much less (Table 1). In complex areas, if map units must be delineated quickly or if data is sparse, it may be necessary to delineate groups of soil classes that commonly occur together as complex map units. Figure c illustrates simple map units of high purity. The legend description of a map unit lists the soil classes it contains, their proportions, and perhaps their pattern.

The map with its legend and a set of class definitions comprise the conventionale pre-computer SIS. The soil properties at any site would be predicted by

- using the map to identify the map unit at the site
- using the legend to ascertain the soil class (e.g.) in that map unit
- using the soil class description(s) to ascertain the properties of that class.

Without a computer this is probably the best that can be achieved. Ideally there will be a one-for-one correspondence between map units and their suitability for some kind of land-use (e.g. Land Use I in Fig. 2). If so, it will be possible to interpret the soil map directly as a land suitability map, just by relabelling the map units and removing boundaries between map units of equal suitability (Fig. 4).

Note, however, that map units are rarely pure (e.g. Table 1) though the 'impurities' do not necessarily behave differently from the classes listed in their map unit description. Similarly there may not be a one-for-one correspondence between map units and soil suitability classes (e.g. Land Use II and III in Fig. 2). Also, once a soil class is substituted for the exact values of the soil properties at a site, it will be possible to predict the soil

Table 1. The purities and uniformities of five map units on two experimental maps in the Netherlands. From: Marsman & de Gruijter, 1982.

Soil phase		Soil suitability classes for:						Purity of map unit[3]
		Rye			Grass			
map unit 1/50 000 map		1	2	3	1	2	3	
A[1]	Soil class	0	*100*	0	*100*	0	0	9
	Map unit[2]	*53*	44	6	*92*	5	3	
B	Soil class	*100*	0	0	*100*	0	0	63
	Map unit	*88*	10	2	*84*	16	0	
1 : 10 000 map								
C	Soil class	0	*54*	46	*54*	46	0	
	Map unit	38	*43*	19	*80*	20	0	6
D	Soil class	*52*	48	0	*100*	0	0	12
	Map unit	*56*	39	5	*91*	9	0	
E	Soil class	*100*	0	0	*100*	0	0	24
	Map unit	*100*	0	0	*98*	0	0	

[1] A-E Substituted for full names.
[2] As mapped by free survey.
[3] Percentage of map unit correctly predicted by its description in map legend.

properties at that site only as lying somewhere within the ranges given in the corresponding class description. More seriously, perhaps, once a soil class name has been substituted for the values of a range of soil properties at a site, it will not be possible to re-assess the site in terms of a later classification.

For a real example, Table 1 lists the proportion of sites within each map unit that are assessed in Suitability Classes 1, 2 or 3 for rye or grass cultivation and the proportion of sites on the dominant soil class in each map unit that are assessed in Suitability Classes 1, 2 and 3. It shows the following.

- Sites belonging to the same soil class usually show a greater uniformity in their suitability class than sites belonging to the same map unit.
- Some soil classes or map units show a class correspondence to a suitability class whereas others do not.
- The suitabilities of a set of soil classes for one purpose need not be in the same order as their suitabilities for another purpose.

Soil mapping

Before discussing the new possibilities offered by the computer, we should first consider what is actually involved in soil mapping. The main procedures fall into three groups, though they may be combined in various ways.

Grid survey

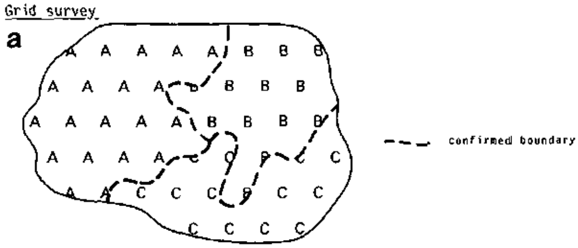
Here the mapper identifies the soil class at each of an array of sites, and then inserts soil boundaries between sites of dissimilar class (Fig. 3a). The sites are usually on a regular grid, since this is more efficient than distributing them randomly. In a pure-grid procedure this interpolation of boundaries is quite mechanical and draws on no other information. If he can, of course, the mapper will use the evidence of changes in vegetation or land surface to position them more precisely.

The spacing of the grid points must be close enough to allow interpolation, which usually means more than 100 observations per square kilometre, according to the complexity of the landscape, so grid survey tends to be used for very detailed or large-scale surveys, or in landscapes where soil differences have no external expression as changes, for instance, in vegetation.

Physiographic of air-photo surveys

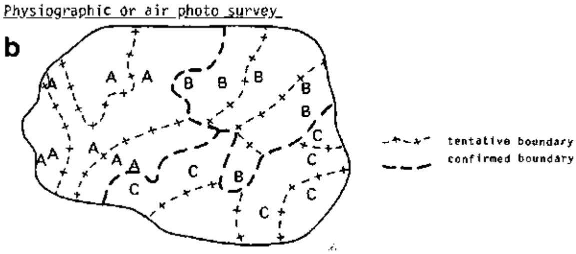
Here, by contrast, the surveyor starts by observing, if not recording all possible discontinuities of soil surface, vegetation and land surface, or previously determined boundaries of geology, etc., that might possibly correspond to the differences in soil class that he wishes to distinguish. He may look for these boundaries on air photographs. He may generate classes of spectral signatures from remote-sensor data by cluster analysis and map their

Grid survey



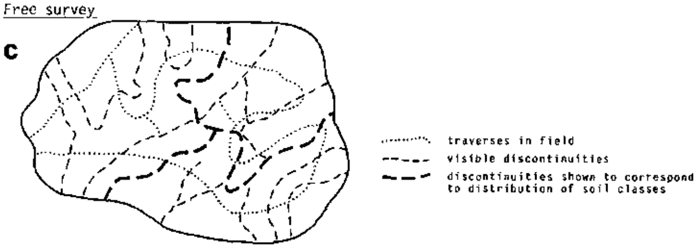
Sites preselected, soil classes identified
boundaries then inserted without intuition or flair

Physiographic or air photo survey



Start by noting all possible boundaries
(i.e. discontinuities in the landscape)
and then deploy field work to discover which of them correspond
to distribution of soil classes and which must be discarded

Free survey



Use of continually unfolding landscape / soil model -
checked by soil surveyor as it develops

Fig. 3. Three types of soil survey procedures.
a. Grid survey
b. Physiographic or air photo survey
c. Free survey

distribution, to position their boundaries. Having established a set of discontinuities as tentative boundaries, the surveyor then plans his field work to establish which of them separate the soil classes he wishes to separate and which do not. The latter are rejected (Fig. 3b); the former may perhaps be refined in detail.

Usually this procedure is employed in surveys with only limited opportunity for soil observations, perhaps less than $20/km^2$, and hence for small-scale maps. It requires that the external expression of the soil boundaries be reasonably consistent.

Free survey

This procedure makes the greatest use of the surveyor's experience and skill. He deploys his field effort on a series of loose traverses and proceeds by continuously developing hypotheses on the relation between the soil boundaries he wishes to map, and the landscape features he can see. His soil observations are mainly to check the soil-landscape models he thinks he perceives, and once confirmed, he uses these to fix soil boundaries (Fig. 3c). In effect, he is using a form of personal discriminative analysis to ascertain the landscape features that identify soil boundaries, but is prepared to use different discriminants in each field. Usually free survey requires some 10-50 observations per square kilometre.

These brief summaries emphasize one crucial point - that it is only in grid surveys, which requires relatively close observations, that soil boundaries may be positioned only on point data. When the density of observations is less than $100/km^2$, one needs information from perceived or inferred discontinuities in the landscape to draw soil boundaries.

Computer-based soil information systems

The computer can of course store soil boundaries (Fig. 5c) or soil maps (Fig. 5e) and descriptions of soil classes or map units, and perform all the operations required in the pre-computer SIS. If there is a one-to-one relation between map unit and soil-suitability class, it can perform the draughtsman's function of re-labelling map units of similar suitability class (Fig. 4a). These are all useful but not new.

What is new, is that it can also store ungeneralized point data, either on soil properties (Fig. 5a) or soil classes (Fig. 5b). Previously the only point-data stored were the class identifications at points on the surveyor's field sheets (Fig. 5g), retained but not generally available. If the detailed point-data is available, it becomes possible to make specific appraisals of all sites for any conceivable land-use, by means of a standard interpretive link or better by means of a discriminative analysis that uses some sites to identify the soil discriminants that determine or conclude with soil suitability for that purpose; the discriminants are used to classify the other sites. Site appraisals ad hoc will usually be more precise and more relevant than appraisals in terms for a single general-purpose classification. This is a big advantage, and has been frequently stated.

What is commonly forgotten, however, is that it is only the sites that are re-appraised (Fig. 4b) ad hoc. *Unless the sites are close*

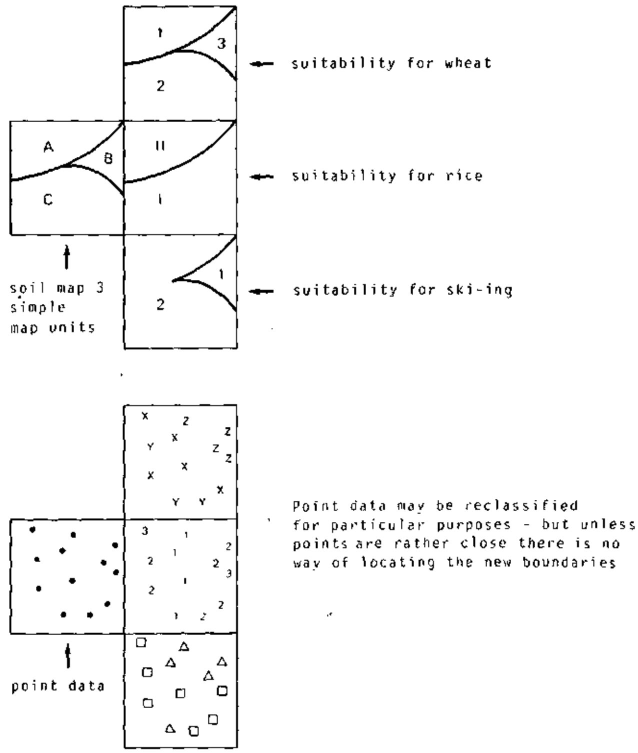


Fig. 4. Map units can be translated into evaluation units if there is one - for - one correspondence.

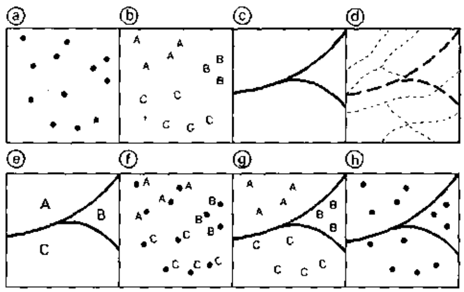


Fig. 5. Soil maps based on four kinds of information (a-d).

enough for grid survey, it will not be possible to delineate the boundaries of the corresponding ad hoc map units (Fig. 4b), because this would require also the kinds of information on tentative boundaries that are employed in our photo or free soil survey. If computer-based SIS is to have the required capacity for interpolating appraisals or classifications ad hoc, it must store as many tentative boundaries as possible in some way or other (Fig. 4a) in the hope that some of them will serve to fix the boundaries of new map units. Figure 6 illustrates this. Figure 6a shows the original general-purpose soil map. The sites for which detailed information was stored were then reclassified ad hoc (Fig. 6b) into Classes D and E which show little correspondence to the general-purpose Soil Classes A, B or C. However all the tentative boundaries, that had been rejected when preparing the original map, had been stored (Fig. 6b) and some of them correspond to the distribution of D and E points closely enough to serve as likely boundaries to Map Units D and E, until it is possible to check them on the ground.

Conclusion

The solution offered here is crude. Nevertheless it emphasizes the real problem. If the ability of the computer to reappraise or reclassify sites ad hoc for any new purpose is to be exploited, we must find some way of fixing new boundaries ad hoc between the sites that can be used when the site density is less than $100/km^2$.

Added later

Giltrap's paper in this volume offers another solution to the same problem. When the data for a site is put into the system, he interpolates values of all significant soil properties at points on a relatively close grid, using kriging separately within each set of plausible physiographic discontinuities, and stores the interpolated points. This may be more efficient than my proposal which would store the discontinuities, and make interpolations as required, though there may be problems where there is a very large number of plausible discontinuities.

Reference

Marsman, B.A. & J.J. de Gruijter, 1982. Kwaliteit van bodemkaarten; een vergelijking van karteringsmethoden in een zandgebied. Rep. No. 1714, Stichting voor Bodemkartering Wageningen, Netherlands.

MIDGE, a microcomputer soil information system

D.J. Giltrap, Soil Bureau, Department of Scientific and Industrial Research, Private Bag, Lower Hutt, New Zealand

Summary

MIDGE (map image display generator) is a microcomputer-based map display and manipulation system which gives a powerful data retrieval system for soil or other land-based information. Because it is implemented on a low-cost microcomputer (currently an LSI 11/2) MIDGE does not require access to large computer facilities or central databases.

MIDGE accesses maps stored on floppy discs in a run-length coded grid format. These maps may be displayed at any scale on a colour VDU (Intercolor 8001G). The storage format allows storage of quite detailed maps in files of 10 kbytes or less so that a large number of single-factor maps of any area can be stored on a single disc.

Maps may be overlaid to produce composites and this allows sophisticated land interpretation. As an example, maps showing predicted pastoral productivity, and a suitability classification based on seasonal production patterns have been generated from basic soil and climatic data. Any users of the system can create interpretive classifications to meet their own requirements and generate the resulting maps from basic land information.

The basic maps used by MIDGE should ideally be produced by interpolation from point data but there is no technical problem in using data from other sources (e.g. digitized soil maps). In general, however, single-factor maps generalized from soil maps appear to be less detailed than similar maps interpolated directly from point observations. Conventional soil maps can also be produced by interpolation and can be displayed with MIDGE.

MIDGE allows a series of maps of the same area with same legend to be stacked in a single file. This facility can be used to store distribution maps for different soil classes or to store means, minima and maxima (based on interpolation errors) for particular soil properties. This gives the user access to information not usually available at all from conventional soil maps.

Because MIDGE is a self-contained low-cost system, it can be used directly by land-information users (e.g. farmers, land planners). These users then have access to a very much larger body of information than could be displayed on a single hard-copy map, how ever sophisticated the information system used to produce that map may

have been.

Introduction

Modern techniques of processing soil survey data and of land evaluation have generated a need for more powerful cartographic information systems than can be provided by traditional maps. A major advance in soil survey technology is the capture and subsequent retrieval of basic soil properties (as determined at a single site) for land evaluation. This avoids the potentially serious information losses associated with taxonomic generalization and allows much more detailed land evaluation.

The simplest technique for storing and mapping individual soil properties is the grid survey (Rudeforth, 1975). In this technique, survey observations are at the intersections of a regular grid. Maps may then be drawn on a line printer or by other means (e.g. grid to vector conversion) based on the actual soil properties observed at each point. An alternative technique is the polygon survey where the landscape is subdivided into a number of polygons taken to be uniform and the properties of the soils occurring in each unit are recorded. This technique resembles conventional soil mapping and the polygon format is used in a number of land information systems (Van Berkel & Eyles, 1981 ; Tomlinson, 1967). In practice, however, it is often impossible to subdivide a landscape into a manageable number of polygons without considerable generalization (i.e. the polygons are markedly varied in at least some properties). Similarly a grid survey often gives an adequate degree of detail only where a large number of grid cells are used thus imposing an unrealistic burden on the soil surveyor who must characterize the soils at each grid cell.

These problems may be overcome by interpolation procedures (de Gruijter & Bie, 1975). With these procedures, survey observations may be at random sites (or at the surveyor's convenience) and the observed properties are then interpolated mathematically to 'target points' on an arbitrarily fine grid. The interpolated grid data may then be mapped in exactly the same way as for a grid survey. A particular interpolation technique that has recently attracted interest in soil science is 'kriging' (Burgess & Webster, 1980; Yost et al., 1982; Giltrap, 1983a). This method gives a theoretically optimum (minimum standard error) solution and also gives a separate estimate of standard error for each interpolated point. Kriging requires substantially more computer time than other interpolation techniques but it has been shown (Giltrap, 1981) that it can be applied routinely to soil surveys for a comprehensive range of soil properties with computer costs that are small compared to total survey costs. With this approach, large numbers of single-factor soil maps may be routinely generated by soil survey.

As soil survey organizations have acquired the capacity to generate (and retrieve) large collections of soil data, there has been an increase in the information requirements for modern land-use management and planning. Land evaluations are required to be increasingly specific and related to particular land-uses. Ideally land evaluation maps should give quantitative predictions of yields of particular crops and responses to particular management practices. Such maps should also give some indication of their reliability and of the entent of land variability over small areas. The resulting demands on a land information system cannot be met by traditional printed maps.

The 'cartographic' information discussed above cannot easily be handled by standard database management systems, which are designed to store and retrieve information about discrete enumerated objects. The 'object domain' of cartographic information is a continuous set of points on the earth's surface and, retrieval requests will often relate to points other than those at which data has been collected.

It is possible to make cartographic data fit the conventional database concept by arbitrarily dividing the landscape into polygonal regions and then treating these regions as discrete objects. This approach, however, leads to problems of cartographic generalization and ensuing information loss. Furthermore geographic relationships between polygons (e.g. contiguity) may be poorly handled and separate programs are often needed for graphical output or for specifying requested search areas (which may not correspond to polygon boundaries).

Alternatively the landscape may be represented by a regular array of grid points. In order to achieve reasonable resolution, however, a large number of grid points will be needed. This leads to database management problems if we attempt to use individual grid cells as the basic entities in a database. It is also very unlikely that a user would wish to retrieve data on the individual grid cell (e.g. obtain listing of all grid cells with attributes in a specified range). The grid cells will, in fact, be used only to represent a continuous geographic surface with output presented either as a map or as a summary of total areas possessing particular combinations of attributes.

We may more appropriately treat the complete arrays of grid cell attribute values as internal computer representations of the spatial distribution of attributes across a continuous surface and treat these distributions as the basic entities in a database. O'Callaghan & Graetz (1981) refer to this as image-based storage. In practice, image-based systems lead to small simple databases (albeit with large complex records) and database management systems turn out to be unnecessary.

The MIDGE system

The MIDGE (map image display generator) system was developed at this laboratory to provide a map display and manipulation system capable of exploiting the amount of data produced by the application of geostatistical techniques to soil surveys (Giltrap, 1983b) and of using modern predictive models to construct land evaluation maps from basic soil properties. MIDGE provides interactive map display on a low-cost colour VDU and is designed to operate on stand-alone microcomputers. The full map-processing capabilities of the system are thus available to any land-use decision maker for a modest capital outlay. This contrasts with systems implemented on large central computers, where the end-user of the land information has access only to some form of computer printout (e.g. computer-plotted map).

MIDGE is currently implemented on an LSI 11/2 microcomputer with 64 kbytes RAM and dual 500 kbyte floppy disc drives using an Intecolour 8001G VDU as the terminal console. As the programs are written in FORTRAN, it should be relatively easy to implement MIDGE on any other microcomputer of comparable capacity. The individual map images are stored as run-length encoded grids in floppy disc files. The run-length encoded format gives substantial data compression and quite detailed map images (200 × 200 pixels or more) can be stored in files of less than 10 kbytes. A single floppy disc can thus hold many thematic map images for any given area.

The MIDGE display is formed from colour patterns, which are formed by printing a character in one of eight possible colours on a different background colour. This restricts the resolution of the display to that of the normal character on the 8001G screen (48 rows by 80 columns) but makes it easy to identify unambiguously a large number (> 50) of distinct units. This contrasts with both vector format displays and conventional image display systems, which give greater resolution but poorer 'readability' for maps with many units. In practice, the limited resolution is not serious as MIDGE allows 'zoom' and 'pan' functions to look at specific parts of the image in greater detail. The maximum detail that the MIDGE display can resolve is limited only by the resolution of the stored image. This is typically about 200 × 200 pixels but much greater detail (1000 × 1000) may be used where the quality of the basic data warrants it.

The MIDGE display is constructed from the stored image by mapping each character position in a fixed (37 × 49) window on the screen to the closest corresponding stored image pixel and drawing the colour pattern of the appropriate unit. The remainder of the screen is used to display the map legend and title, the current display scale and a statistical summary of unit areas. The display is manipulated by a set of single-letter commands with possible numeric arguments (e.g. the command S10000 sets the display scale to 1 : 10 000). These commands are supplemented by simple keystroke operations for cursor movement across the map display and colour pattern definition.

MIDGE files may contain several map images of a single land surface provided that these images share a common legend. This helps avoid file proliferation by keeping groups of related images together.

Examples of groups of images that could be stored together in this way include
- maps of a single soil property (e.g. pH) measured at different depths;
- maps of predicted cash return for a range of alternative land uses;
- maps of seasonal (or monthly) soil moisture deficit (or any other temporally varying land property;
- percentage distribution maps for soil classes or vegetation types.

A simple one letter command is available to step (or jump) through the images stored in a multi-image file.

Probably the most powerful facility MIDGE possesses, however, is the ability to overlay two or more images to produce a new composite image. For simple overlaying, the composite map is produced pixel by pixel by classifying the combinations of units in the source images for the appropriate pixel. MIDGE can also produce overlaid images from source images defined on different grids from each other or from the composite image. This facility can be used to 'compose' images of neighbouring areas into a single large image.

The classification of the combination of source image units is interactive. The computer displays the combinations actually found and the user keys in the cõmposite unit numbers. The user is then prompted for the new map title and unit descriptors and the composite map image is then displayed and stored. This approach means that no provision need be made for combinations of source units which do not occur in practice.

The applications of map overlay extend far beyond the simple exercise of simultaneously displaying more than one image. Because the combinations are reclassified, a wide range of predictive models may be used to generate land evaluation maps from basic land properties. Several examples of this will be given in the following section. The classification may be used with a single-source image to alter isarithm intervals. An initial isarithmic map may be stored with very fine intervals (up to 255 classes) and these classes may be grouped as required.

Examples of use

MIDGE has been used to display map images generated by geostatistical techniques from current soil surveys. The overlay technique has been used to generate predictive land evaluation maps based on soil and climatic properties.

In a survey of the urban environs of Masterton, described by Giltrap (1981), a wide range of basic soil properties (e.g. soil thickness, depth to particular layers, soil colour and texture at specified depth intervals, were kriged from observed points to a regular grid. These grids were transferred to the MIDGE system for display and manipulation. In this area, the most important properties for assessment of agricultural potential were thought to be available moisture storage capacity and drainage. Available water storage was estimated from the thickness of the various layers within the root zone and the particle size distribution of these layers. Drainage was represented by the average abundance of orange mottles in the root zone. Because these images were produced by kriging, it was possible to produce error maps for them and mean, minimum (mean - s.d.) and maximum (mean + s.d.) images were produced for each property and stored in multi-image files.

These water storage capacity and mottle abundance images were overlaid and the resultant classes reclassified to give a composite map in which the soil properties were expressed as land limitations (none, slight, moderate or severe). The resulting map resembled conventional land-suitability maps in general format. Good visual separation of the classes was obtained with a total of 16 units. The composite map was easily produced and any required classification of the units could have been used.

Water balance calculations were performed for the various available water classes and used as inputs to a range of predictive models. A simple model estimated the (water-limited) summer grass production from the ratio of actual to potential evapotranspiration. An image showing this prediction was generated by reclassifying the available water map. In this survey climate was constant; has a larger area been involved, available water would have been overlaid with appropriate climatic maps. The resulting image gave an apparently good general representation of the pattern of summer productivity and showed that summer production over the landscape varied by a factor of about 2.

A more sophisticated model may consider production for all four seasons. Evidence based on production trials (unpublished work) suggests that pasture production in any season can be predicted from radiation, soil and air temperature and actual evapotranspiration. A pastoral suitability classification has been set up on the basis of these relationships. This classification takes account of the seasonal variation and annual reliability of production as well as the total mean annual potential production. A map image of these classes was constructed on MIDGE by combining the available water storage with the relevant climatic data.

MIDGE has also been used to display and manipulate images of conventional soil maps. The simplest way to do this is to digitize a soil map and convert from vector to grid format. In a recent soil survey, geostatistical techniques were used to generate soil maps from point observations. These techniques led to an interpolated grid array with probability estimates for each possible soil class at each grid cell. A soil map was generated from this grid by simply allocating each grid cell to the class with highest probability (i.e. simple legend). Alternatively, had particular combinations of probabilities occurred over extensive areas, these areas could have been mapped to a complex unit.

With MIDGE, however, it is possible to store the complete probability map for each class in a single multi-image file. This enables the user to display the (probable) distribution for any class (or combination of classes). In particular, minor classes that are never dominant and hence are not shown on a conventional map can be displayed as required. The same technique could presumably also be used for other complex nominal data (e.g. vegetation types).

Conclusions

A microcomputer-based system for map-image display and manipulation has been developed. This system has been shown to be suitable for sophisticated land evaluation with large amount of information derived from geostatistical processing of soil survey data. The data compression achieved by run-length encoding in the image files means that microcomputer systems (using floppy discs) have adequate capacity to hold all the information likely to be required for any given application. Because the map image is treated as the basic entity of data, large database management systems are not required and hence the system is not dependent on links to larger computers.

Because the image is interactively generated, a very large number of different images are available to the user. This means that new techniques for displaying map information are available. One example of this that has been demonstrated is the use of multi-image files to store distribution maps for each possible soil class as opposed to the traditional soil map, which shows only the dominant class at each point. Another example is the display of statistical ranges (mean, minimum and maximum) for isarithmic maps rather than just the mean.

The generation of new images by overlay has been demonstrated and the reclassification of overlaid units has been used to generate land-evaluation maps based on predictive performance models. This opens the way to sophisticated land-use planning as more detailed land information and better models of how this data relates to land performance become available. Even where such information is not available, however, MIDGE or other similar systems can contribute to the better use of whatever data is available by providing a clear simple display of land properties to the end-user of this information. Because of the low (and falling) cost of microcomputer systems, MIDGE can be made available to many land-information users in developing or developed countries and not just a few large institutions.

References

Burgess, T.M. & R. Webster, 1980. Optimal interpolation and isarithmic mapping. I. The semi-variogram and punctual kriging. J. Soil Sci. 31: 315-331.

Gruijter, J.J. de & S.W. Bie, 1975. A discrete approach to automated mapping of multivariate systems. Proc. Int. Cartos. Ass. Com. III. Enschede. p. 17-28

Giltrap, D.J., 1981. A practical application of spatial interpolation for the rapid production of a wide range of special purpose soil maps. In: Information systems for soil related data. Proceedings of the second Australian meeting of the ISSS workshop on soil information systems, Canberra, Australia, 19-20 February 1980. (Eds A.W. Moore, B.G. Cook & L.G. Lynch). Pudoc, Wageningen. pp. 73-82.
Giltrap, D.J., 1983a. Computer production of soil maps. I. Production of grid maps by interpolation. Geoderma 29: 295-311.
Giltrap, D.J., 1983b. Computer production of soil maps. II. Interactive map display. Geoderma 29: 313-325.
O'Callaghan, J.F. & R.D. Graetz, 1981. Software for image-based information systems: its application to integrating Landsat imagery and land tenure data for rangeland management. In: Information systems for soil and related data. Proceedings of the second Australian meeting of the ISSS workshop on soil information systems. Canberra, Australia, 10-20 February 1980. (Eds A.W. Moore, B.G. Cook & L.G. Lynch). Pudoc, Wageningen, pp. 64-72.
Rudeforth, C.C., 1975. Storing and processing data for soil and land use capability surveys. J. Soil Sci. 26: 155-168.
Tomlinson, R.F., 1967. An introduction to the geo-information system of the Canada land inventory. Canadian Dept. of Forestry and Rural Development. 23 pp.
Van Berkel, P.R. & G.O. Eyles, 1981. The New Zealand resource inventory database for the national soil and water conservation organisation. In: Information systems for soil and related data. Proceedings of the second Australian meeting of the ISSS workshop on soil information systems. Canberra, Australia, 19-20 February 1980. (Eds A.W. Moore, B.G. Cook & L.G. Lynch). Pudoc, Wageningen. pp. 104-121.
Yost, R.S., G. Uehara & R.L. Fox, 1982. Geostatistical analysis of soil chemical properties of large land areas. II. Kriging. Soil Sci. Soc. Am. J. 46: 1033-1037.

A probabilistic approach to mapping soil boundaries

T.M. Burgess, Schlumberger Cambridge Research, P.O. Box 153, Cambridge, England
R. Webster, Rothamsted Experimental Station, Harpenden, Herts. AL5 2JQ, England

Summary

Normal procedure for soil survey in many parts of the world is to observe the soil at regular intervals on transects, to classify the soil at each site, and to draw boundaries between sites of different soil type. In all such surveys, there is the problem of identifying all the boundaries without excessive sampling. This problem can be solved if the statistical distribution of distances between boundaries is known and a reasonable risk of missing boundaries can be accepted. In many instances, the distribution appears to be a negative exponential and it is then necessary to know only the average distance between boundaries in order to plan a survey with known risks. If the same exponential distribution applies to all classes of soil in a region, the optimal scheme is a strict grid whose interval is determined from the mean boundary spacing and the declared risk. The paper outlines the principles and illustrates the procedure.

Introduction

Soil maps showing regions divided into small parcels, each of a particular type of soil continue to be in demand, especially for land development. The techniques for making such maps by free survey and aerial photo-interpretation depend on identifying relations between soil and physiography, vegetation and perhaps existing land-use. Where such relations do not exist or are obscure, mapping must be based solely on observations of the soil itself in pits and boreholes. Boundaries can be placed only between points where the soil has been observed to be of different kinds. Beckett and Bie (1975) showed that in these circumstances, the effort required for mapping is related to the frequency of boundaries. If there are changes in soil between neighbouring observations where the soil is judged to be the same, they will not be detected and the resulting map will be deficient.

The only sure way of identifying all soil boundaries would be by observing the soil everywhere; that is clearly impractical. So in practice, some boundaries must be missed. Several questions arise in consequence. Can we say how many boundaries were are likely to omit by sampling in a particular way? For a given sampling effort, is there a configuration of observations that will incur least risk of omitting boundaries and will in that sense be optimal? Alternatively, if we can specify a certain risk of missing boundaries, i.e. the maximum proportion of missed boundaries that we are prepared to tolerate, can we plan a strategy that will just meet that tolerance? If so, that again will be an optimal strategy.

The answer to all three questions is 'yes', provided the appropriate distribution functions are known. In this paper, we discuss briefly the nature of these functions and their generality, how they can be estimated, and how they can be used to plan a survey making the best use of resources.

The distribution of boundary spacings

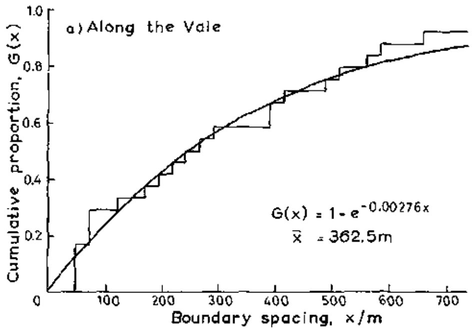
Consider a region in which there is a pattern of soil types and that transects are laid accross the region. These transects will intersect the boundaries between the soil types. The distances between successive intersections can be measured, and the resulting set of values will have a frequency distribution that may be regarded as the distribution of soil boundary spacings. Of greater interest in the present context is the cumulative distribution. Suppose that of N interboundary distances, $n(x)$ are less than some distance x. Then $n(x)/N$, the proportion of spacings less than x, estimates the probability that the distance between two succesive boundaries is less than x. Let this probability be $G(x)$, which can also be used to denote the cumulative distribution function of boundary spacings. Later, we shall require the probability density function, which is the first derivative of $G(x)$, and we shall denote it by $g(x) = \mathrm{d}G(x)/\mathrm{d}x$.
Figure 1 shows two examples of the stepped function for one of the clay vales of the English Midlands. Also plotted on each graph is a smooth curve that fits the observed distribution rather well. It is the graph of the negative exponential

$$G(x) = 1 - e^{-\underline{\lambda}x} \tag{1}$$

It is the single parameter $\underline{\lambda}$, which is the average intensity of the distribution and has the dimension $\mathbf{L}^{-1}$. The quantity $1/\underline{\lambda}$ is the mean distance between boundaries. On a transect of length L, the number of boundaries crossed has a Poisson distribution with parameter $\underline{\lambda}L$. We therefore have in this instance a simple mathematical description of the distribution of spacings between soil boundaries. As it happens, the parcels of different types of soil in this region cover much the same area on average. Elsewhere there are parcels of contrasting size, and in some instances shape. In these instances, one will want a separate distribution function for each type of soil and, in our experience (Burgess, 1979), these too are often negatively exponential. So, despite its simplicity, this sort of distribution seems sufficiently widespread to consider general sampling strategies based upon it.

Determining the risk

Let us return to the situation described in the introduction and illustrated in Figure 2. Suppose an observer examines the soil at



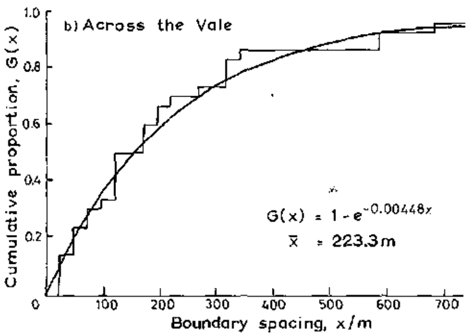


Fig. 1. Cumulative distributions of soil boundary spacings in the clay vale near Stow-on-the-Wold, England. a. Along the vale. b. Across the vale. The smooth curves are those of the negative exponential with λ equal to the reciprocal of the mean boundary spacing, $\bar{x}$.

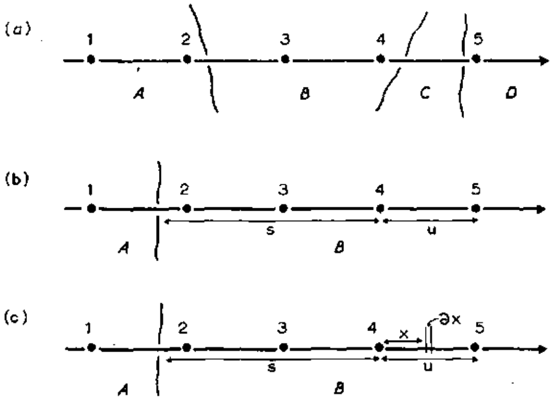


Fig. 2. Diagram showing mapping errors and sampling decisions.

Points 1, 2, ... along a transect that crosses three soil boundaries. He would correctly identify the boundary between soil types *A* and *B* between Points 2 and 3. He would not identify soil type *C* because he would cross two boundaries. So by sampling at this frequency he would miss one boundary. If, however, he were to double the sampling intensity, he would have an additional observation between Points 4 and 5 and would detect both boundaries. He would, of course, be doubling the cost of observation, and in many parts of the transect to no advatange, since there are no boundaries to detect, e.g. between Points 1 and 2, and between Points 3 and 4.

We may regard such a sampling procedure as perfect when a transect segment crosses just one boundary, in error if it crosses more than one, and extravagant if it crosses none. Some compromise must be struck between the last two, and this introduces the concept of a risk. We can define the risk of a sampling scheme in this context as the probability of committing an error, that is, of crossing two or more soil boundaries between successive sampling points on a transect. The aim in planning a scheme then becomes one of choosing sampling intervals as large as possible for a pre-determined and acceptable risk.

The situation at any stage in the survey becomes as illustrated in Figure 2b. The observer has detected a boundary between Points 1 and 2 and has made his latest observation at Point 4, a distance s from the boundary. He wants to make his next observation at a Point 5, a distance u from Point 4 in such a way that the probability of crossing two boundaries equals the acceptable risk. Intuitively he may feel that this probability depends in some way on s, the distance he has already travelled since crossing the last boundary. So the distribution functions that he is concerned with appear to be conditional on s, and may be designated $G(x|s)$ and its derivative $g(x|s) = dG(x|s)/dx$, where $G(x|s)$ is the probability that the distance between successive boundaries is less than $x + s$, knowing that it is greater than s.

It is one of the general laws of probability that for any two events, A and B, the conditional probability of A, knowing B, is given by

$$P\{A|B\} = P\{A \text{ and } B\}/P\{B\} \tag{2}$$

So if X is a distance between two successive boundaries then

$$G(x|s) = P\{X \leqslant x + s | X > s\} \tag{3}$$

From Equation 2, this is equivalent to

$$\begin{aligned} G(x|s) &= P\{X \leqslant x + s \text{ and } X > s\}/P\{X > s\} \\ &= P\{s < X \leqslant x + s\}/\left[1 - P\{X \leqslant s\}\right] \\ &= \left[G(x + s) - G(s)\right] / \left[1 - G(s)\right] \end{aligned} \tag{4}$$

Its derivative, the probability density, is then

$$g(x|s) = \frac{g(x + s)}{1 + G(s)} \tag{5}$$

Consider now the situation shown in Figure 2c. The surveyor has reached Point 4 and is planning to site his next observation at Point 5, a distance u from Point 4 (P_4). For any distance x less than u, the probability of crossing a boundary within the interval (x to $x + \partial x$) is $g(x|s)\partial x$. If a first boundary does indeed occur in that interval, the probability of at least one more boundary occurring in the interval ($P_4 + x$ to $P_4 + u$) is $G(u - x)$. The risk of crossing two or more boundaries is then given by multiplying these probabilities and integrating over all possible values of x, from 0 to u, thus

$$\text{Risk} = \int_0^u g(x|s)\ G(u - x)\,\mathrm{d}x \tag{6}$$

The distribution functions used in Equation 6 may be either general ones for all classes of soil or specific to one class in which the last observed point lies, as appropriate.

Risk with negative exponential distribution

Let us now consider the risks for the negative exponential distribution. The cumulative distribution, $G(x)$ was given in Equation 1. Its derivative is

$$g(x) = \underline{\lambda}\mathrm{e}^{-x\lambda} \tag{7}$$

If we substitute in Equation 5, we discover that $g(x|s)$ also equals $\underline{\lambda}\mathrm{e}^{-\underline{\lambda}x}$, and therefore $g(x|s) = g(x)$. Inserting the right side of Equation 7 into Equation 6 gives

$$\begin{aligned}\text{Risk} &= \int_0^u \underline{\lambda}\mathrm{e}^{-\underline{\lambda}x}\ \{1 - \mathrm{e}^{-\lambda(u - x)}\}\mathrm{d}x \\ &= 1 - \mathrm{e}^{-\underline{\lambda}u}(1 + \lambda u)\end{aligned} \tag{8}$$

This result means that where boundary spacings have a negative exponential distribution, the risk depends only on u and not on s. No matter how far the observer has travelled from the last boundary, the risk depends only on the length of the forthcoming sampling interval. So for constant risk, the sampling interval must be constant. The optimum sampling scheme can then be planned on a simple regular grid.

If the distribution functions change from class to class, sampling is performed on separate grids within each soil parcel. This is a less tidy operation, and a surveyor might choose this solution only where the expected boundary spacings were very varied.

It remains only to determine the sampling interval for such a grid. This can be done by plotting Equation 8 with risk as a function of $\underline{\lambda}u$ (Fig. 3). For given risk, the desired interval is read from the graph in terms of $\underline{\lambda}u$. The value of λ, the reciprocal of the mean

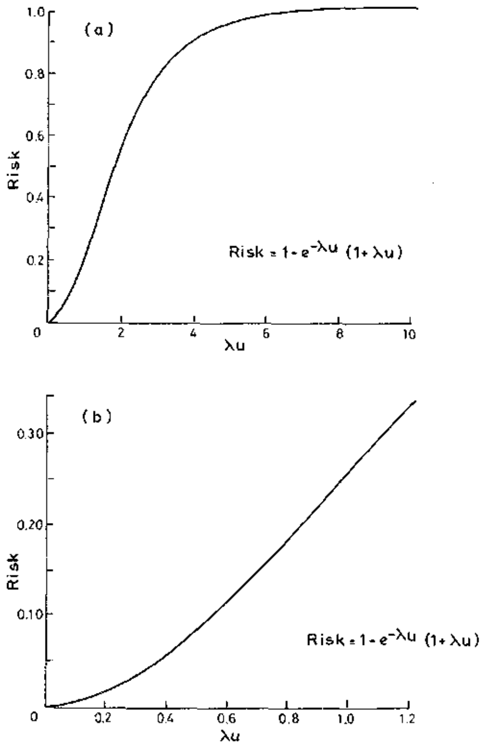


Fig. 3. Risk graphs for negatively exponential distribution, a. Risks up to 1. b. Risks up to 0.3.

boundary spacing, x, will have been estimated by reconnaissance, so that u is readily calculated. Usually, a surveyor will wish to operate with a small risk, so the graph need not be extended beyond a risk of about 0.3, as in Figure 3b. From this graph, it is seen that the optimum interval is about 0.52 $\bar{x}$ for a risk of 0.1, and about 0.81 $\bar{x}$ for a risk of 0.2.

For the example of the clay vale, these intervals are 187 and 296 m, respectively, along the vale and 116 m and 181 m across it. So the optimum survey of this region with a risk of 0.1 is a rectangular grid with an interval of 116 m in transects aligned across the vale and a spacing of 187 m between transects.

General discussion and conclusions

The strategy we have outlined above applies to the situation where the soil boundary spacings have a negatively exponential distribution and are much the same for all classes. We have indicated the refinements that can be made where there are different distributions for different classes. Nevertheless, this requirement is practical only if the soil class can be identified in the field. If laboratory measurements or the judgment of a person more experienced than the one who records the observations are needed to determine the class of soil, the simple procedure is the best that can be applied. So there are many circumstances in soil surveys for land development where a regular grid of observations is best.

The negative exponential can be considered as a special case of a more general distribution, the gamma distribution. In our experience, this applies widely (Burgess, 1979). Where identification of soil type has to await laboratory measurement or assessment, the gamma distribution also leads to a regular grid with an interval dependent on the mean boundary spacing. However regular sampling is not optimum if soil classes can be identified in the field, since the expectation of crossing a boundary depends on the distance travelled from the previous one. This may accord with intuition, but it requires an elaboration of the sampling strategy that is beyond the scope of this paper.

References

Beckett, P.H.T. & S.W. Bie, 1975. Reconnaissance for soil survey. I. Pre-survey estimates of the density of soil boundaries necessary to produce pure mapping units. Journal of Soil Science 26: 144-154.

Burgess, T.M., 1979. Optimizing procedures for soil survey at large to medium scales. D. Phil. thesis, Oxford University.

Dutch soil survey goes into quality control

B.A. Marsman and J.J. de Gruyter, Netherlands Soil Survey Institute, P.O. Box 98, 6700 AB Wageningen, the Netherlands

Introduction

Today, many map users are no longer interested in pedological data or in general descriptive information about mapping units and soil properties. Users want quantitative information on soil properties related to the particular purpose for which the map is being used. The need for quantitative information increases as computerized data is being used at a wider scale, e.g. by users working on land-consolidation schemes. Such users need information on relevant properties of soils actually present in a given mapping unit. Until now, little information of that kind was available.

For the future, it will be a challenge for Soil Survey Institutes to collect and provide reliable data on the variability in soil properties of mapping units and their composition. This can be realized by creating an information system of quality data. Such a system will provide data for adapting elements of survey procedures (e.g. observation density, map scale, content of the legend) to specify circumstances. It can also be used for land evaluation. Besides, the system enables us to present information on the advantages and disadvantages of using soil maps for specific uses.

Previous work on map quality in the Netherlands

The first steps in studying map quality were undertaken in 1972 when the Netherlands Soil Survey Institute participated in the work on soil variability and map quality at Oxford (Beckett & Burrough, 1971; Beckett & Webster, 1971). In this project, the usefulness and efficiency of different survey procedures were tested. Our contribution to this project consisted of producing free survey maps on small, medium and large scale. Purity of soil units in the free survey maps on medium and large scale ranged from 55-75 % (Beckett, 1975).

It became clear that investigations on quality of soil maps are very important for soil surveys. Problems and difficulties in such extensive investigations come to light as well. The purity of only a few soil delineations of simple mapping units in sand and river-clay deposits was investigated.

The results showed lower purities on the subgroup level than for similar data on purity in other countries. Sometimes very low purities were achieved at the phase level of the Dutch soil classification system. It seemed, however, that for much of the impurities, the behaviour of (or response to) agricultural crops was similar to that of the dominant soil profile. Purity was then an incorrect measure for quality, accuracy or usefulness (Wilding, 1965).

Studies on quality of survey methods

In 1975, the Dutch soil survey started an extensive investigation on the quality of soil maps produced by different survey procedures in a sandy area near Laren in Guelderland. Survey procedures differed in one or more of the following aspects: 1. delineation of mapping units (field or by the proximal method of Thiessen polygons); 2. the choice of observation points (free or random; 3. degree of generalization.

Maps were tested by a stratified random sample of test borings. Besides various purity measures, standard deviations of soil properties within mapping units and their homogeneity in terms of suitability classes were used as quality measures (Marsman & de Gruyter, in preparation).

We give some results for two of the methods, as applied on the scale 1 : 10 000:
- free survey map, made by free choice of observation points and delineation in the field;
- proximal map, made by random choice of observation points and proximal delineation.
Fragments of both maps are shown in Figure 1. Just as in previous studies, purity on phase level was very low for both maps. Still, mapping units were found to be reasonably homogeneous in terms of suitability classes for grass and winter rye.

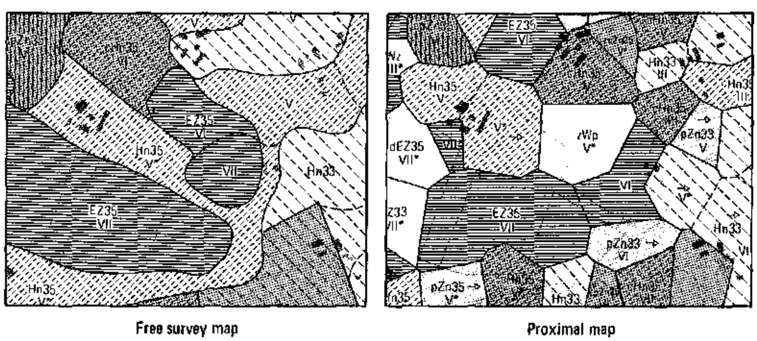


Fig. 1. Fragments of soil maps, produced by two different survey methods. Soil map near Laren (Gld.), scale 1:10 000, observation density 1.2 per ha. Free survey map: free choice of observation points abd field delineation. Proximal map: random choice of observation points and proximal delineation.

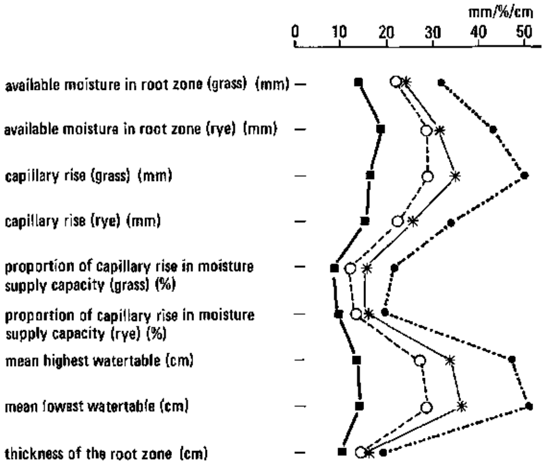


Fig. 2. Standard deviations of 9 soil properties within the:
o------o mapping units of the free survey map
——— mapping units of the proximal map
■———■ taxonomic units of the applied classification system
●-·-·-·● surveyed area

Standard deviations for 9 soil properties are shown in Figure 2. Represented are the standard deviations within mapping units of both maps, within the taxonomic units of the applied classification system and within the entire surveyed area. Standard deviations within mapping units are higher than those within classification units, but lower than those for the surveyed area.

Compared with data from England (Beckett & Webster, 1971; Beckett & Burrough, 1971), we found lower purities and, for most properties, lower standard deviations within mapping units. In contrast to data from other countries, standard deviations within taxonomic units are distinctly lower than those within the mapping units. Expressed as a percentage of the standard deviations within the surveyed area, standard deviations within mapping units range 55-85 %, as compared with 25-55 % for the taxonomic units. In our opinion, both the lower purity of soil maps and standard deviations within taxonomic units are predominantly caused by the great detail in the Dutch classification system. In this classification, overlapping classes are seldom used and criteria are often strictly morphometric.

As shown in Figure 2, standard deviations are lower for the free survey map than those for the proximal map. Expressed as a proportion of standard deviations within the surveyed area, they range from 56-79 % for the free survey map and between 70 and 83 % for the proximal map.

Besides differences in standard deviations, there are other differ-

Table 1. Number of mapping units and delineations and the average area per delineation for a free survey map and a proximal map.

	Mapping units	Delineations	Average area per dealineation (ha)
Free survey map	52	145	2.5
Proximal map	148	356	1.0

ences between the two maps. As shown in Figure 1, the proximal map shows a fragmented pattern with many very small delineations. The average size of soil delineations on the free survey map is 2½ times as large as on the proximal map. The number of mapping units (3 times) and the number of delineations (2½ times) are much higher on the proximal map (Table 1). This reduces the legibility of the map and its usefulness. Delineation in the field on scale 1 : 10 000 produced delineations of satisfactory size coupled with lower variability of soil properties. This procedure yielded better results than the proximal procedure.

Recent developments

Investigations on quality of soil maps are very expensive and time-consuming. So they have mostly been carried out by research departments while survey departments have rarely been involved. Application of the results to soil survey is limited so far, because they rarely solve the types of practical problem arising in routine soil surveys.

In order to overcome these operational problems, we have searched for a method that is simple and cheap enough to be incorporated in standard soil survey practice. Random transect methods seem promising, because on average only a short time is needed to locate an observation point in the field and to go to the next one. Random transect methods were used by Powell & Springer (1965) and by Steers & Hajek (1979). The one we used was constructed for estimating composition and homogeneity of mapping units with special attention to cost-effectiveness. The procedure was tested on two mapping units of a soil map on the scale 1 : 10 000.

At the first stage, some soil delineations were randomly drawn with replacement and with probabilities proportional to their size. At the second stage, two points were randomly marked in any of the selected delineations, being the starting points of a transect in N-S and in E-W direction, respectively.

The observation points on each of these transects, including starting point, were those 25 m apart on the straight line with the given direction, within the given delineation. By previous division of the largest delineations, the length of the transects was restricted to a mximum of 300 m. The position of selected delineations and transects is shown in Figure 3.

Fig. 3. Location of transects in the delineations of two mapping units on the soil map of Lievelde, scale 1 : 10 000.

Profile descriptions and other observations are made at each observation point (Fig. 4). For mapping unit Hn53-VI 7, delineations were selected, with 14 transects and 61 observation points. For mapping unit Hn53-IV, 6 delineations were selected, with 12 transects and 67 observation points. Four additional observations are

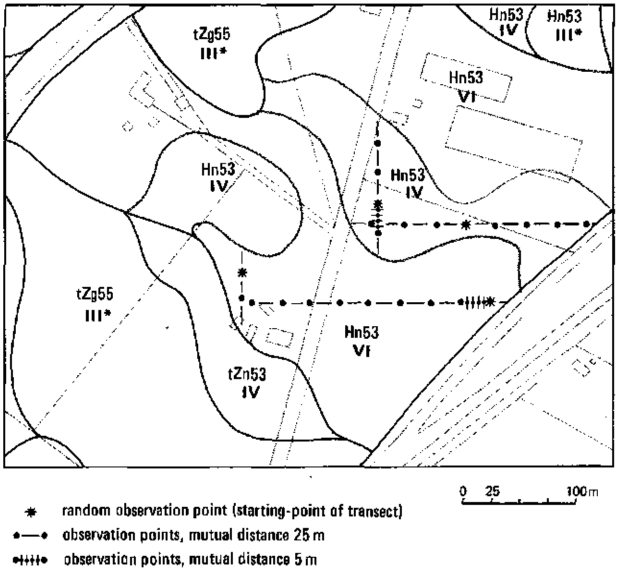


Fig. 4. Detail of soil map Lievelde with transects in two delineations

made at intervals of 5 m in each of the selected delineations. They were between the randomly selected starting point and the next observation point 25 m away.

The data collected on transects will be used to establish quality of mapping units on soil maps. Besides expressions for purity and composition of mapping units, estimates of the mean and variability of soil properties and differential characteristics can be made. Information about spatial variability over short distances will be obtained from the observations at intervals of 5 m. This will be

Table 2. Confidence intervals (90 %) for mean and standard deviation of three properties within two mapping units of the soil map Lievelde, scale 1 : 10 000.

		Mapping units	
		Hn53 - VI	Hn53 - VI
Mean highest watertable (GHG)	mean:	50.8± 5.8	31.4x14.4
	standard deviation:	14.2± 2.1	15.4± 2.2
Mean lowest watertable (GLG)	mean:	134.7±16.1	120.8±21.0
	standard deviation:	24.5± 3.7	23.4± 3.4
Thickness of the root zone	mean:	43.6± 6.1	42.1± 4.5
	standard deviation:	12.0± 1.8	10.8± 1.5

used to evaluate the efficiency of the intervals of 25 m. Some results for both mapping units are given in Table 2. Confidence intervals (90 %) for mean and standard deviation of three soil properties indicate that the accuracy of the data obtained is sufficient. Moreover, statistical estimation is relatively simple and the extent of additional sampling does not prohibit incorporation of this procedure in standard soil surveys.

Discussion

Laren-type studies on map quality, although expensive and time-consuming, will probably have little impact on soil survey practice. Survey departments are usually not involved in such studies and the results often do not meet the specific demands of surveyors executing a particular soil survey. Map-quality investigations should therefore be directed towards measures related to usefulness of soil maps, such as variability of special soil properties, land qualities and homogeneity for suitability classes. Data on strict pedological features such as hue, value and chroma of mottles can have scientific value but, in themselves, contribute little to the usefulness of soil maps. Purity of mapping units appears to have limited value as a measure of map quality, because it is defined on the basis of soil characteristics with varied influence on soil behaviour. The latter is of primary interest.

Introduction and development of systematic quality control of soil surveys requires cooperation between different departments within the Survey organization. This implies a step-by-step approach, each step providing a useful piece of information. A random transect method seems promising for collecting useful data. The purpose is to create an information system with a growing amount of data on quality of particular mapping units, characterized by their pedological composition and the variability of relevant soil properties. Such a system will be essential to provide information to map users on advantages and disadvantages of maps for different purposes. The data are also important for internal use for adapting survey procedures and on improving land evaluation.

References

Beckett, P.H.T., 1975. The cost-effectiveness of terrain evaluation. Vol. 2: Final Technical Report: Review of Field Work 1971-74. Department of Agricultural Science, University of Oxford.

Beckett, P.H.T. & P.A. Burrough, 1971. The relation between cost and utility in soil survey. IV. Comparison of the utilities of soil maps produced by different survey procedures and to different scales. J. Soil Sci 22(4): 466-480.

Beckett, P.H.T. & R. Webster, 1971. Soil variability: a review. Soils and Fertilizers 34(1): 1-15.

Bakker, H. de & J. Schelling, 1966. Systeem van bodemclassificatie voor Nederland. De hogere niveaus. Pudoc, Wageningen.

Marsman, B.A. & J.J. de Gruyter, in preparation. Quality of soil maps; a comparison of survey methods in a sandy area. Soil

Survey Papers, no. 15, Netherlands. Soil Survey Institute, Wageningen.
Powell, J.C. & M.E. Springer, 1965. Composition and Precision of Classification of Several Mapping Units of the Appling, Cecil and Lloyd Series in Walton County, Georgia. Proc. Soil Sci. Soc. Am. 29(4): 454-458.
Steers, C.A. & B.F. Hajek, 1979. Determination of Map Unit Composition by a Random Selection of Transects. Soil Sci. Soc. Am. J. 43(1): 156-160.
Voort, W.J.M. van der, 1981. De bodemkundige variatie binnen twee enkelvoudige legenda-eenheden (Rn44C and Rd90A) op de bodemkaart van Nederland, schaal 1 : 50 000. Boor en Spade 20.
Wilding, L.P., R.B. Jones & G.M. Schafer, 1965. Variation of soil morphological properties within Miami, Celina and Crosby mapping units in West-Central Ohio. Proc. Soil Sci. Soc. Am. 29(6): 711-717.

Computer-aided soil cartography: some recent Franco-British projects

David P. Bickmore Associates, Laurel Farmhouse, Bletchingdon, Oxford OX5 3DP, England

The main stream of soil cartography all over the world still takes for granted the manual processes of traditional cartography. The preparation of beautiful coloured maps in large formats and displaying 50-100 different soil categories each in a different (and hopefully) recognizable colour is typified by the production within the last 6 months of soil maps on 1 : 250 000 covering Scotland (Soil Survey of Scotland, Macaulay Institute) and England & Wales (Soil Survey of England & Wales, Rothamstead Experimental Station). Both these operations have been carried out entirely by manual methods: they have obviously been expensive to produce although there has until recently been little evidence that computerized production would have cost no more money, whatever the other advantages it might provide.

But in my own experience, computerized soil mapping has a 15-year history. One of the first examples (ECU 1971) was the digitization of the Soil map of the Abingdon area of England on 1 : 63 360 and the production of a colour map on that scale of 'Predicted underdrainage treatment for arable land use'. This was an experimental exercise in digitizing a manual compilation of soil boundaries and replaying them regrouped and in finished form on a flat bed plotter; colour tints were produced by a largely unautomated method. The problems of this first experiment lay mainly in the difficulties - at that time - of obtaining accurate digital results - or rather in the costs of correcting errors or changes of detail or changes of classification within the digital system. The possibilities of using the data to play back other information on different scales were also limited by manually produced base maps on which shape and position were only relatively the same as between scales. Thus the 'correct' soil lines in their digital form would have to be altered to conform to the manually generated patterns of topography on smaller-scale base maps which - right or wrong - were 'different'. The use of digital data for soil patterns but not for the base map detail also accentuated the lack of a relationship within the computer between, e.g., contour levels or stream patterns and the soil data. In 'soil information system' terms this is unfortunate and now, we would argue, is unnecessary.

A subsequent ECU experiment was conducted with the Kansas Geological Survey and led to the publication of the Wakarusa Quadrangle, Kansas Atlas in 1974. This was based on the digitizing of soil, topographic and geological patterns on scales of 1 : 20 000 and 1 : 48 000 respectively. These patterns were then edited interactively on a graphic display so that something approaching a coherent data base became available. The published atlas includes some 12 colour maps (on 1 : 24 000 and 1 : 48 000), directly derived from the three sets of data. At this stage, the work was still arguably more expensive than by manual methods, largely because of the costs of interactive editing (of which there was plenty) for editing the pattern lines. The costs of producing colour printing originals by traditional cut and peel processes showed no saving over manual methods despite plotting outlines on the tint areas quickly and economically from the edited database in the computer. At this stage too, the raster process of fast areal overlay within the computer was not available and the facility of relating any particular element of topography to another of soil was slow and expensive.

In 1981-1982, the Geothematic Department of IGN Paris collaborated with INRA and the Chambres d'Agriculture du Cher et du Loire-et-Cher in the preparation by digital methods of several *Cartes des Sols de la Région Centre* - this time using raster rather than vector methods. This has resulted in sets of maps in which a number of elements have been combined on 1 : 50 000 and are displayed seperately on 1 : 100 000. Thus soil classifications are shown on 1 : 50 000 in colours, with the nature of substrata, depth and texture shown by annotations: the texture of the profile, obstruction of drainage, and porosity of rock are each separately identified. From such a base of information, many derived maps are possible - and, this time, are claimed to be substantially cheaper than by classical methods. Part of the economics of the French system lie in the automation of the cut and peel process that also virtually eliminates colour proof corrections - unless, of course, the soil surveyor changes his mind or his data.

My slides illustrate one of the set of these maps covering the 1 : 50 000 carte des sols sheet Aubigny-sur-Nère and with plots at 1 : 100 000 for the same area of agricultural suitability, of constraints related to excessive water, of texture, and mineral reserves - all these aspects being directly - and automatically derived from the Carte du Sol itself. These maps are concerned with quite complex situations but have been designed for printing in 4 colours - a further economy of significance. IGN claims that the total cost of producing each set is less than producing the single 1 : 50 000 sheet alone by traditional methods.

The occasion of the Euro-Carto I meeting in Oxford at the end of 1981 exposed a number of us to the new IGN raster techniques and led directly to the first of a series of Franco-British experiments this was the Eco/Topo experiment in the Dolgellau area of Wales and produced a map at the scale of 1 : 50 000. The background to this experiment lay in an approach to ecological mapping by use of topographic and pedological data, both of which were available: the first objective of the experiment was to have digital records of the data that could be related to each other rapidly both by statistical and cartographic methods - e.g the areas of particular types of soil above a particular altitude. The second objective was to produce a digital map at about 10 % of the costs of an earlier digital mapping experiment carried out between ECO and the British Nature Conservancy Council. The third objective was to derive vegetation categories from soil ones (see last paragraph). And the fourth was to see how detail from a normal 1 : 50 000 topographical map responded to treatment by raster methods. These I now describe.

The materials assembled for this initial experiment and supplied to IGN Paris consisted largely (Items i - vii) of separations of the Ordnance Survey (OS) map in the forms that happen to exist under their classical system of cartography; this was supplemented by particular compilations (vi and vii).

(i) Rivers and coastline. The line positive without tone filling was used. Names, lake outlines and double-line rivers were deleted from this document leaving only single-line rivers and streams; gaps in the river pattern (e.g. under bridges) were joined up. In addition the water tint positive was photographically thickened to give a solid version of areas of sea, lakes and double-line rivers.

(ii) Roads. The OS map uses colour fillings in red, brown and yellow and these were photographically combined to produce a line pattern of main roads. (Note that the cased roads on the map are combined in black with houses, names and much other 'noisy' detail making them slow and expensive to prepare for scanning.)

(iii) Built-up areas. Again the OS map outlines these on its complex black plate along with cased roads etc.: however a pink tint filling for built-ups exists as a separation and this identified all built-up areas except isolated small houses, barns, etc.

(iv) Woods. The green tint (again thickened to a solid) was used rather than the outline which was yet another element in the black 'culture' drawing.

(v) Contours. Selected height contours for a part of the area were rescribed to eliminate gaps (e.g. for values). The rescribing was partly done on to two separate originals, one showing coastline, 500, 1000, 1500, 2000 feet - and the other showing the intermediate levels of 250, 750, 1250 and 1750 feet. (In fact, these could as easily have been derived from a negative of the OS original.) The intention here was to derive slope zones automatically.

(vi) Soil boundaries. These existed as a compilation at 1/50 000 (intended originally for representation by classical methods on the 1/250 000 soil map). An overlay was prepared showing a key number for the soil class of each polygon. For the area concerned, there were some 30 separate soil classes distributed over 137 polygons: a table was prepared that indicated how the soil classes were to be grouped into ecological classes.

(vii) Parish boundaries. These are no longer shown on OS map at

1 : 50 000 scale. However a compilation of them was made and a separate overlay to this allocated each parish an individual key number.

The scanning operation was carried out at two resolutions - 16 and 8 pixels per millimetre. Higher or lower resolutions could have been used and the choice was conditioned partly by scanning time, but even more with a view to the advantages of decreasing the amount of digital data produced. A resolution of 1/8 mm at 1 : 50 000 produces individual pixels of about 6 m on the ground: while this lower resolution is graphically satisfactory for areas - woods, soil polygons, etc. - it produces some 'saw-tooth' edges on some roads which the higher resolution used for rivers etc. does not display - at least to the naked eye. It should be noted that the total time for the scanning was just under 4 h. The soil compilation has used rather thick lines and some interactive touching up (c. 1 h) was desirable to help produce the fine-line skeleton plot in raster form.

The scanning treated all the shapes and patterns as areas - the line data (e.g. rivers, polygons, roads) are long thin areas and the system uses software to reduce these to a skeleton form. The system can translate a scan of areas like lakes into a line of pixels exactly along the edge of the black and white on the original document.

The next stage of the process is conducted on a separate Vax computer and consists of organizing the - so far - white pixels within any polygon into sets (3 min). A plot of the soil polygons was then mounted on a digitizing table and each polygon was 'touched' roughly in the middle so as to allocate the set of pixels involved to a particular attribute code. This process of allocating attributes by touching is quick (47 min in this case, i.e. 20 s per polygon); it is done twice so as to ensure accuracy of attributes per polygon and hence virtually to eliminate those expensive colour proof errors that often bedevil the equivalent process in classical cartography. Typically this check identified 3 areas with no codes and 2 small unclosed polygons.

The soil compilation - because originally intended for a 1 : 250 000 map - had differences in detailed shape from the 1 : 50 000 topography - e.g. along the coastline. These discrepancies were corrected interactively (both for the boundaries and the fillings) - and took about 30 min. The soil data set was then ready for automatic production of a set of screen positives. These can use conventual screens or ruling patterns or can explore more exotic new designs at minimal cost. In this case, 8 vegetation categories were derived from the 30 soil categories held in the database.

The topographic details such as rivers, roads and contours, were treated broadly on the same manner, but the separation of single-line rivers from double ones was of course preserved in the database and, for example, suggests one logical way of selecting 'main rivers' for smaller scale maps. For roads, the scanned original - after skeletonizing - furnished a digital centre line: almost any design of road symbol can then be derived - in this case red cased lines which were combined with detail of built-up area. It would also have been possible to have plotted a fine black outline round the edges of the built-up areas had that seemed graphically desirable. Note that the graphics to be used are kept separate from the information about shapes.

One of the great advantages of a raster system is its ability to overlay areas and derive statistics as well as graphics. The areas of different soils in each parish were listed automatically. It would have been equally simple to have measured, for example, the lengths of single-line streams per soil category.

No experiments were made on this project for handling alpha-numerics. However you will see on the French 1 : 100 map (Aubigny - Aptitudes Agricoles) that key letters are shown in each soil category: these letters were automatically generated by the system from the data held about the polygons and were then plotted roughly in the middle of each parcel. This not only eliminates another tedious manual stage but has much better chance of being accurate since it is simply an alternative manifestation of the stored data about the attributes of each parcal: the raster plotter - suitably instructed - does the rest. In the Dolgellau experiments, the key numbers were added manually.

One of the claims for a database is that the information in it can be readily and regularly updated. The IGN system certainly permits a patch of revised soil data to be compiled, scanned and then to replace the equivalent area of the older version. Unless the manual compilation of the new area has smoothed the junction of lines between the new patch and the original map, the edge matching will have to be done interactively (see paragraph above on soil compilation). Of course, once the revision has been made to the information in the database, it is there for all scales of map and for many other derivative processes.

Much of the speed and elegance of the system developed by IGN Paris depends on the use of raster methods of digitizing and of plotting. For some purposes, however, it may be desirable to hold the data on magnetic tape in vector as opposed to raster form. This raster/vector translation was not used in the Dolgellau experiment but was performed for a recent 1 : 50 000 geological map. The data on the tape describe the shapes of the bounding lines of polygons between nodes and furnish the attributes that lie to the left and right of the line. Such tapes could, for example, then be used for replotting the pattern lines of the map on a flat bed or drum plotter, if that were desirable. For some purposes too, a vectorized format is a convenient and more economic way of holding digital data.

The total cost of the Dolgellau experiment - up to colour proof stage - amounted to c. FF 15 000, a figure that did indeed imply a

10-fold reduction in the costs per square kilometre of a previous digital experiment, also on 1 : 50 000 scale in this area with broadly the same objectives. It is difficult to press cost comparisons of this kind too far for the actual map materials and compilations that represent the starting point of any particular operation may vary substantially: one is regrettably far from the logical solution of ensuring that compilations are prepared in a form that is optimum for the new processes of map making. Nevertheless the experiment does give grounds for optimism in this aspect of digital cartography - an aspect that has too long remained discouraging.

There are theoretical arguments about processing environmental data where it is collected, as opposed to bringing it back to some central site for processing. Cartography has been an increasingly centralized activity over the last century and the prospects of using fast and cheap computer processing to decentralize it have many attractions. It has however to be said that at present raster scanning systems are expensive (around £500 000) to buy and install: their staff need to acquire particular skills and to develop specialized software. And these facts point to centrality at least for the processes of mass digitizing. Once that 'weeping wall' has been surmounted, the digital data really can be accessible to those who wish to do their own things in their own ways. In IGN's case, their 50 or so years of human labour in technical experience in raster cartography is an asset even more valuable than the sophisticated system that they use.

Many countries have at last begun to address the formidable problems of organizing national cartographic databases, or geographical information systems, as opposed to using digitizing equipment simply for map making - a latter-day equivalent of scribing. This paper is not the occasion for a discussion of the nature of databases for soil and topography. But raster cartography has some relevance to this subject and it would be sad if cartographic opinion continued to lack experience in these techniques. There is no doubt about use of raster methods in the whole field of remote sensing: bridges between that and digital cartography need building and with some urgency. By 1985, the French satellite - SPOT - will be providing data over the land surface of the world at a resolution of 10 m, in three wavebands, and at a time repeatability of roughly once a month. In relation to this imminent flow of raster data about the environment, one must ask how manually based soil cartography is to develop.

It has to be added - almost as a postcript - that the ecological or vegetation patterns in the Dolgellau area that were derived from the soil data did not agree particularly well with a manuscript vegetation map of the same area drafted a decade or so earlier. Part of the differences were in terms of the classifications used. What this comparison has to say about deriving vegetation patterns from soil and vice versa lies outside my professional scope. This seems a field in which skills in developing remote-sensed data may contribute in a significant way. What is, I hope, beyond doubt is that digital cartography offers techniques that may be important for soil scientists.

Use of computer processing in production of 1 : 50 000 and 1 : 25 000 pedological maps

W. Eckelmann & K.-H. Oelkers, Soil Survey of Lower Saxony, Postfach 510153, Stilleweg 2, D-3000 Hannover 51, Federal Republic of Germany

Summary

During routine production of pedological maps, it has become essential to use computer processing, mainly in order to cope with the vast amount of available archive data and field data during the course of preparation of the map. Since the 1 : 5000 soil-evaluation map and the 1 : 25 000 pedological map are prepared from the same data, they can both be produced by one and the same process. With a flow chart, the various stages of this process are described, including the use of the computer: acquisition and processing of data from external sources; pedological mapping; and preparation of a fair draft map. The method should allow various thematic maps to be produced from existing data.

Introduction

Because of the extremely large amount of data to be dealt with when producing or evaluating pedological base maps, the Survey of Lower Saxony is studying the feasibility of using a data-processing system in a pedological survey of the state.

The use of a data-processing system requires some preparatory work, including the establishment of a data key and form sheets to guarantee the standardized acquisition of data. In addition, one must develop a flow chart to keep track of what has been done when coordinating the expected operations. This chart is central to the following discussion (Fig. 1).

Work-steps fpr pedological mapping by the Soil Survey of Lower Saxony

Pedological mapping in the Soil Survey of Lower Saxony begins with a preparatory phase, in which all available data are perused, evaluated and prepared for use in the compilation of a map (Fig. 1). These data are contained in
- topographic maps, e.g. for information on the relief;
- historical maps, e.g. to demonstrate plaggen soils;
- older soil maps, for the pedological information contained in them;
- maps and soil profiles made from soil assessment data;
- geological maps, especially for the information they contain on the initial soil-forming substrate;
- forestry soil maps for ecological and pedological information.

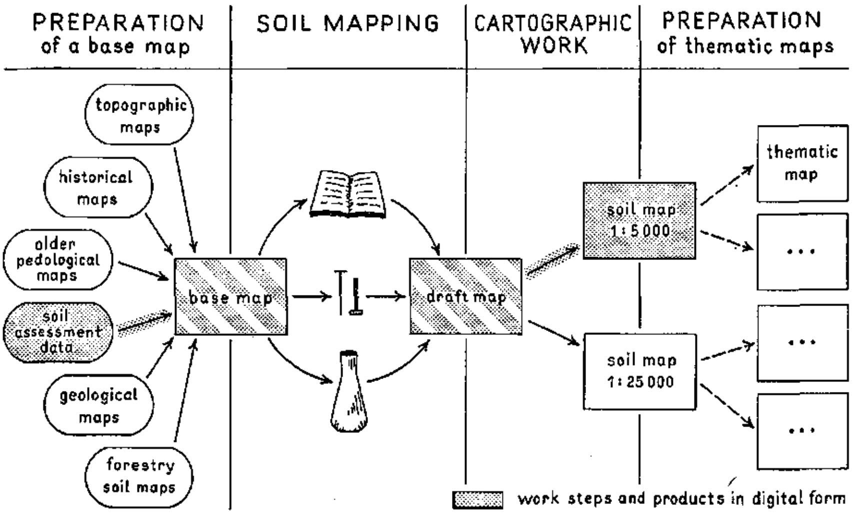


Fig. 1. Flow chart for pedological mapping by the Soil Survey of Lower Saxony. From Oelkers & Vinken (1980).

Information that is essential or useful for pedological mapping is taken from these existing sources. It is compiled into a base map for the field work. The correctness and completeness of this base map are checked in the field. Staff take cores, for which a detailed description is recorded; they analyse soil samples, to back up the field observations, and they revise scientifically all the results obtained.

The condensed and very complex result of archive studies, field observations, soil analysis, and scientific review is a draft map. It is processed into a soil map, scale 1 : 5000 or into a generalized soil map, scale 1 : 25 000.

Computerized operations and products

The operations shown in the flow chart (Fig. 1) have in the past been carried out without the assistance of data processing methods. Because of the vast amounts of data to be processed in the preparatory phase and the insufficient availability of pedological data in the complex soil maps, some of these operations will be computerized. A program, called DASP, has been developed especially for this purpose to handle and evaluate geoscientific alphanumerical data (Kühne, 1983).

Most of the data are from the soil assessment programme started in 1934 covering all agricultural land in Germany. These data are contained in maps on a scale of 1 : 5000 showing soil-assessment boundaries, the site where the soil profile typical for each area was taken, and an assessment of each area. The descriptions of the profiles typical of each area are listed separately in field notebooks. There are about 150 profile descriptions per map, each consisting of 3 or 4 horizon descriptions. Three or more kinds of information are available about each horizon: Thickness, texture, and various characteristics of the horizon. As there are about 10 000 maps to be dealt with in Lower Saxony, this is an abundance of data that can hardly be handled. So computers are needed first to process these data.

Figure 1 also shows operations and products to be computerized:
- preparatory work, e.g. standardization, acquisition and transposition of soil-assessment data;
- plotting of all or part of this data on a base map;
- correction of soil-assessment data before further processing for the final draft map.

Finally, fair drafts are prepared for the production of the soil map, scale 1 : 5000. These operations are dealt with in more detail in the following section.

Preparations for the production of a base map from soil-assessment data: standardization and acquisition

Preparations for production of the base map begin with standardization and acquisition of the following soil-assessment data: sites where the soil profiles typical of the area were taken, the corresponding profile descriptions in the field notebooks, and the area boundaries.

The coordinates of the soil-profile sites are being digitized and stored in a data file together with the profile descriptions. The boundaries of the assessed areas are also being digitized. Since this procedure is time-consuming, it is planned to use automatic acquisition equipment in the future, e.g. a scanner. As the form of the profile descriptions initially made within the framework of soil assessment is not suitable for computerized systems, the acquisition of the profile descriptions requires several operations. The acquisition itself is therefore done together with a certain amount of correcting and supplementing the data to standardize the format.
In a further operation, the data are automatically transposed (or translated), into modern standardized terms (Fleischmann et al., 1979). The necessary pedological terminology has been defined in a 'Datenschlüssel Bodenkunde', a pedological data key (Oelkers, 1983). This key makes it possible to transpose the soil-assessment data automatically into modern terms. The necessary transposition program was developed by a working group at the University of Münster.

The transposed assessment data are stored in files which are accessible to the previously mentioned DASP data-handling system. Using DASP, the data can be systematized, supplemented and retrieved in various forms. The following facilities are available to the user:

```
TK25 3520

PRONUM   KULTUR   BODART   ZUSTAN   WECHSL   ENSTHG   BODENZ
102      A        S        4        0        D        24
ACKERZ   DATUM    KLIMAR   KLIMAK   WASSER
25       1941 5   +4

3.0  Ah   mSfs  h3
2.0  Bv   mSfs  h2
2.0  Bv2  mSfs
Bv2  fSu3
```

Fig. 2. Tabular form of a soil-assessment profile: complete form.

```
*103 lS 4 D            *104 lS 4 D
2  Ah  Sl3 h3          1.5Ah  Sl3 h3
3  Bv2 Sl3 br2         3  Bv  Sl2 br3
2  BvSwSl2 r2          2.5Bv2 Sl2
3  Sw2 mSfSr2          2  BvSwSl3 r3
                       2  Sd  Sl4 r3

*118 lS 4 D            *119 lS 4 D
2  Ah  Sl3 h3          2  Ah  Sl3 h3
2.5Bv  Sl3 br3         4  Bv2 Sl3 br2
2  BvSwSl3 r2          2  BvSwSl3 r2
3.5Sw2 Sl2 r2          2  Sd  Ul3 r3
```

Fig. 3. Tabular form of soil-assessment profiles: abbreviated form.

- a great variety of lists;
- columns giving all the information for a soil profile or selected parts of it;
- representation of selected information in thematic maps with symbols at the sites where the soil profiles were taken.

Forms of representation on the base map

According to the structure of the landscape, three forms of representation are available to the user to portray the soil-assessment data on the base map.
1. The soil profile can be printed in lists, in which they can be arranged according to any given criterion. The position of these profiles can be plotted on a map containing the soil-assessment boundaries. Figure 2 shows a soil profile from such a list. Such a listing takes up quite a lot of space on paper. The method of tabulation shown in Figure 3 is much more efficient in terms of space needed for the print-out. It usually contains quite sufficient soil-assessment data.
2. Is there enough space, the soil profiles can be printed in abbreviated form in the relevant positions together with the sites where the soil profiles were taken on a map of the soil-assessment boundaries (Fig. 4). This type of representation is only suitable for relatively uncomplicated land in which fewer than 100 profiles

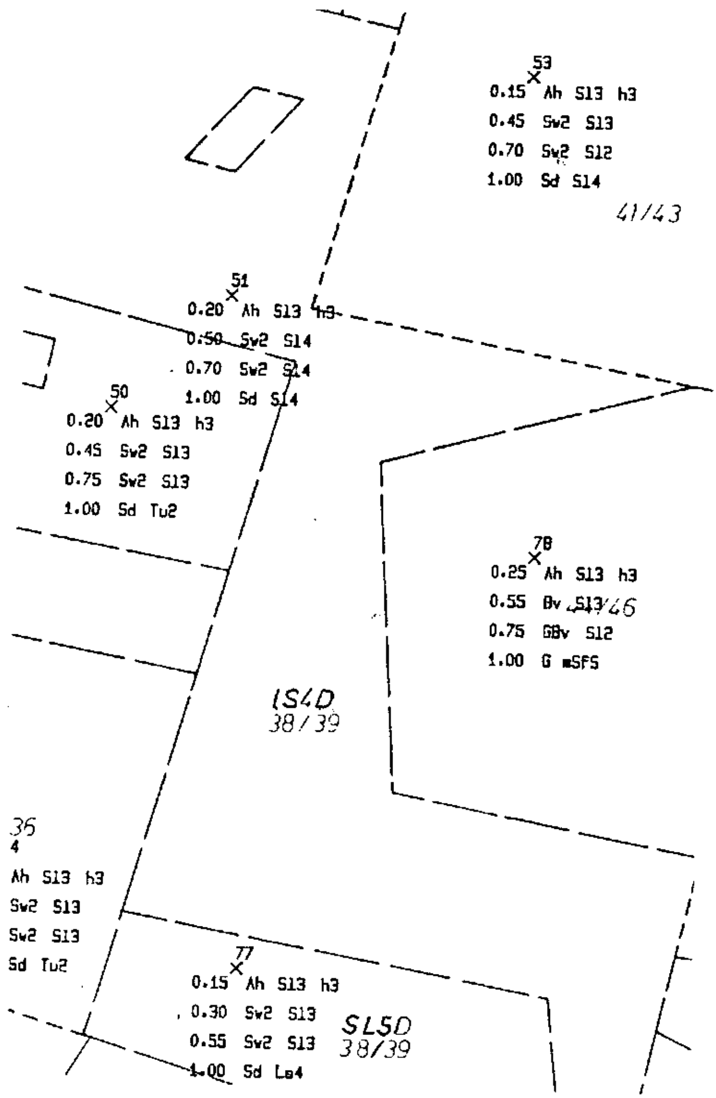


Fig. 4. Soil-assessment profiles plotted on a map containing soil assessment boundaries. From Preuss (1979 and 1980).

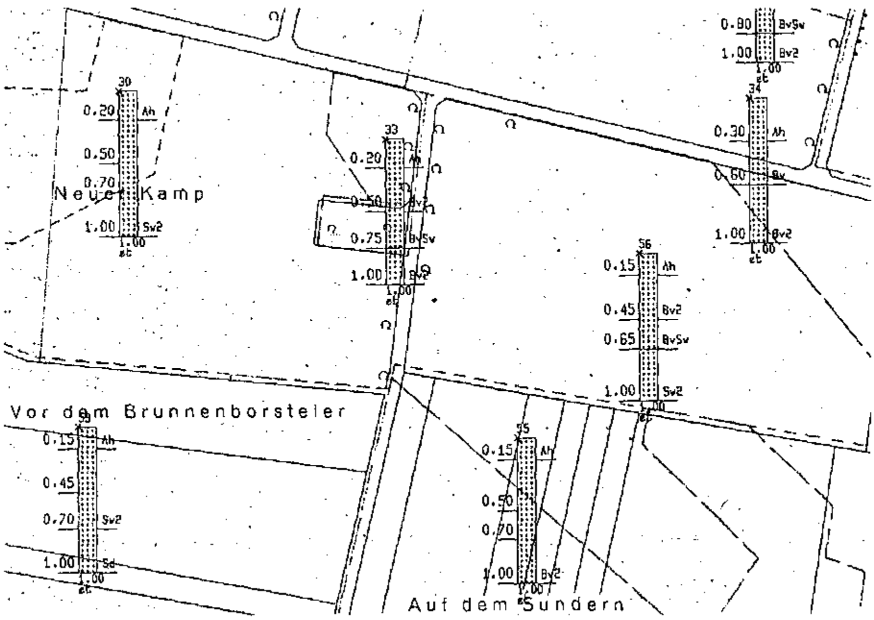


Fig. 5. Columnar form of soil-assessment profiles plotted on a map containing soil-assessment boundaries.

occur in the area of a 1 : 5000 map sheet.
3. Finally, one can print the soil profiles directly on the map in columnar form)Fig. 5). These columns contain the following information:

- at the top, the profile number;
- on the left, the thickness of each horizon;
- on the right, a row of symbols representing the soil horizon, texture, and other relevant information;
- Texture is represented by graphic symbols within the column.

This last type of representation is quite suitable for field work, but it requires more space than shown in Figure 4, in which the profiles are in list form. To avoid overlapping columns, one may plot only selected columns on the map.

The most suitable of these methods of representation can then be selected for the base map. Certain relevant data still have to be added, manually, for instance from the geological map of the area or from historical maps. In this way, the most complete map possible is obtained. With this base map, the actual field work can begin.

During mapping, certain data on the base map (about 10-20 % of it) will turn out to be incorrect. After the field work has been completed, a list of the soil columns is plotted; this has proved to be an ideal basis for checking and correcting the automatically transposed assessment data. The next step is to put the information in the DASP file, including the assessment data, updated by correction where necessary. The corrected assessment data are now available for pedological problems.

The applications of these corrected assessment data include the following:
- preparation of fair drafts for 1 : 5000 soil maps;
- preparation of soil maps on smaller scales than 1 : 5000;
- production of thematic maps;
- soil projects and preparation of expert comment.

Production of fair drafts of soil maps on scales of 1 : 5000 and 1 : 25 000

Currently, only the fair drafts of soil maps on the scale 1 : 5000 can be produced by our computer-based cartographic equipment. The forms of representation just described, or combinations of them, can be used. It is therefore possible to produce the 'Soil Map based on Soil Assessment, 1 : 5000' (DGK 5 B; Mückenhausen & Merten, 1966) by computerized methods. Previously, this map has been produced by conventional methods.

The most suitable presentation for the DGK 5 B is the one with soil-profile columns; this is in fact similar to that used previously for this map (Fig. 6). It has not yet been decided which information should be included at the sides of the column. Horizon thickness, horizon symbol, soil texture, humus and calcium carbonate are possible. So far, due to lack of space, this information has been given in abbreviated form. However, we are working on a program that can print out soil-profile data on the map in text form.

In order to produce a fair draft of a soil map on the reduced scale 1 : 25 000, one must carry out a 'smoothing' operation on the digitized 1 : 5000 soil-assessment data. Data-sorting programs are available to group similar soil profiles according to assessment class or soil type and printing them in list form. The graphic part of this generalizing procedure still has to be carried out by hand. The result is reduced to a scale of 1 : 25 000 photographically.

An important step connected with reduction of the scale from 1 : 5000 to 1 : 25 000 is the merging of several soil types into larger units. These new units are no longer the same as those of the 1 : 5000 maps; thus, the descriptions of the units need revision. It is necessary to develop a system of description which is compatible with computer-based cartography. Such a system for description of soil units must cope with soil units that include a broader spectrum of soil types over large areas. When setting up a universal system for description of soil units, one must bear in mind that subsequent soil maps must as far as possible be directly comparable, and must form a suitable basis for preparation of thematic maps at a later date.

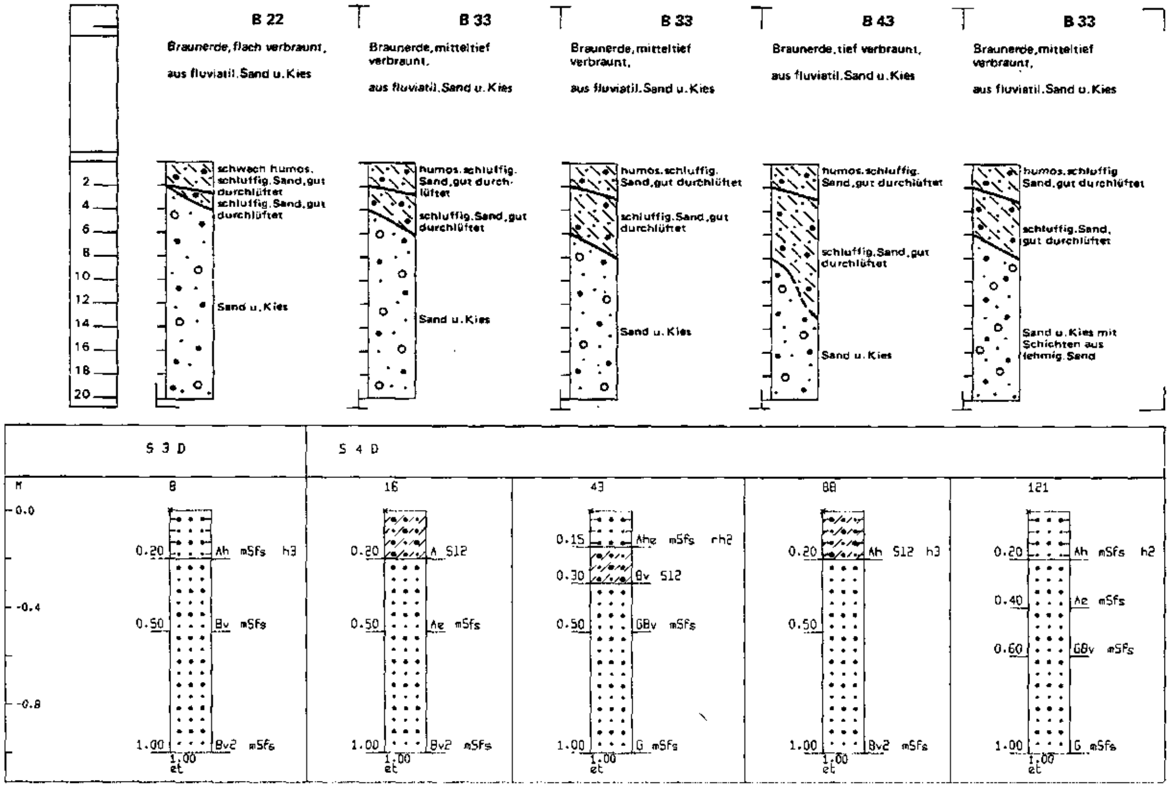


Fig. 6. Columnar form of soil-assessment profiles. Upper half. Manually produced form with legend in words. Lower half. Computer plotted form with abbreviated form of legend.

Production of thematic maps

The chief advantage of producing soil maps with the computer is that draft maps are simpler to make and that one can obtain selected information for the production of thematic maps. Since the 1 : 5000 soil map is stored in digital form, any required profile-site information can be obtained from the database with the wide range of retrieval facilities offered by the DASP program. This information can then be sorted according to any given criteria and plotted on a site map using different symbols. The soil boundaries are then copied onto this map, allowing the site data to be related to the soil distribution (Fig. 7). In this way, areas with the same symbol can be combined, an operation which still has to be carried out manually.

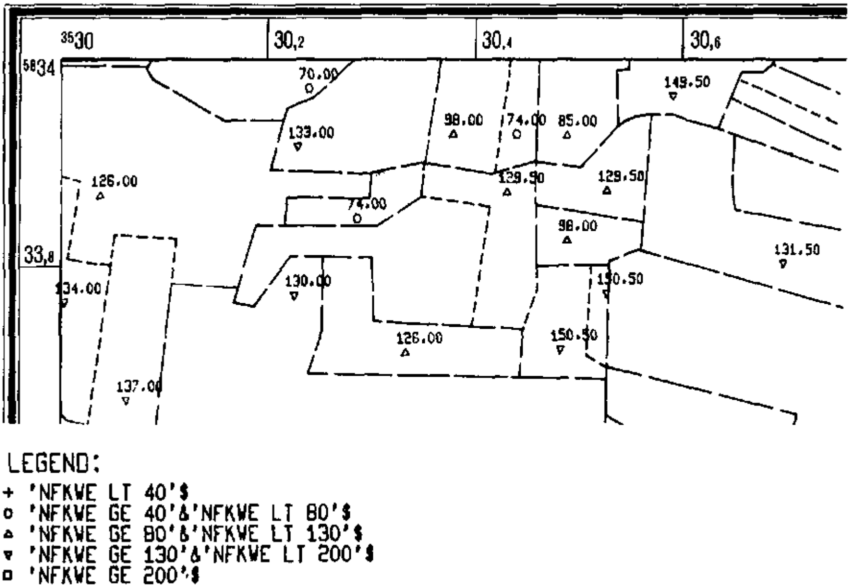


Fig. 7. Thematic map 'Capacity of the Soil for Retention of Plant-Available Water' (NFKWE). Capacity (in mm) is printed at the sites of the soil profiles together with symbols representing various capacity classes (GE = greater than or equal; LT = less than).

We hope we will be able to store digital information not only on the profile sites, but also graphical information related to spatial distribution. This would make it possible to plot only those soil boundaries relevant to the particular thematic map desired. At present, irrelevant soil boundaries have to be removed by hand or photographically. Manipulation of graphic elements requires a special type of databank organization, which we do not yet possess.

Our computer-based cartographic facilities do yet not extend to production of thematic maps on a scale of 1 : 25 000, since the soil map does not exist in digitized form on this scale. It is only a matter of time, however, before this facility exists. A facility already available is computer evaluation for a 1 : 25 000 map on the scale 1 : 5000 and subsequent reduction to the desired scale of 1 : 25 000.

References

Fleischmann, R., E. Hacker & K.-H. Oelkers, 1979. Erfahrungen bei der systematischen Übersetzung der Geländedaten der Bodenschätzung - eine Vorstufe zur automatischen Datenverarbeitung. Geol. Jb. F6: 3-28. Hannover.

Kühne, K., 1983. DASP - Ein System zur Verwaltung und Auswertung geowissenschaftlicher Daten. Geol. Jb., A (in press).

Mückenhausen, E. & H. Merten, 2966. Die Bodenkarte auf der Grundlage der Bodenschätzung. Minist. Ernährung, Landwirtsch. u. Forsten Nordrhein-Westfalen, Düsseldorf. 40 p.

Oelkers, K.-H., 1983. Datenschlüssel Bodenkunde. Geol. Jb. (in press).

Oelkers, K.-H. & R. Vinken, 1980. Möglichkeiten des ADV-Einsatzes in der bodenkundlichen Landesaufnahme. Geol. Jb., F 8: 23-37. Hannover.

Preuss, H., (1979). L4PUSI - Ein Graphikprogramm zur Punktsignaturen-Darstellung. Programmbibliothek von NLfB/BGR, Hannover. 16 p.

Preuss, H., 1980. L4DINO-Programmm zur automatischen Konstruktion von Profilsäulen mit Signaturen an DIN 4220 orientiert. Programmbibliothek von NLfB/BGR, Hannover. 15 p.

Voss, H.-H., 1981. Soil data: Capture, handling, and display using the system DASP. Departement des Sols, Institut National Agronomique, Paris-Grignon, 4: 97-100.

Remote sensing in land and soil monitoring

J. Lamp, Institute of Plant Nutrition and Soil Science, C.A. University, Olshausenstr. 40, D-2300 Kiel, Federal Republic of Germany

Introduction

Increasing numbers of people are facing natural and man-induced destruction and degradation of land potential and soil fertility. Desertification, forest clearing and erosion as well as salinization and sodification, acidification and toxicity, chemical, physical and biological degradations of soil endanger up to 30 % of arable land according to the 'Global 2000 Report'. Thus, survey and monitoring of these processes is a first step in control of worldwide risks to mankind.

There are hopes that multispectral and multitemporal remote sensing by satellites could help in solving these problems. Conventional stereophoto-interpretation, suited to medium-scale evaluations, and ground-truth surveys have to assist the global approach. Potentials and problems of remote sensing in monitoring land changes, soil degradation and ecology will be demonstrated, as illustrated by the major steps in analysis of data about north Germany: a bitemporal land classification based on Landsat data, soil erosion detection and thermal inertia mapping by airborne scanner data. The results were mainly evaluated by the DIBIAS image-analysis system of the German Aerospace Research Establishment near Munich (DFVLR).

The account begins with a sketch of the principles of radiant emission, platforms and sensors.

Emission, platforms and sensors

Physical bodies emit electromagnetic energy at a rate (M) closely related to their temperature ($M = \underline{\varepsilon} \cdot \underline{\sigma} \cdot T^4$ where $\underline{\varepsilon}$ is emissivity; $\underline{\sigma}$ is a constant, T is thermodynamic temperature. Thus, the sun with a temperature of about 6000 has a strong emission with a maximum wavelength ($\underline{\lambda}$ max) of about 0.5 µm and the earth with a temperature of about 300 has a much smaller emission with $\underline{\lambda}$max of about 12 µm.

The radiation from the sun is absorbed by constituents of the atmosphere (e.g. water, carbon dioxide, ozone or dust). So passive methods of remote sensing are usuable only within given 'windows' (i.e. spectral intervals) of the electromagnetic spectrum (Fig. 1). The human eye covers the range 0.4 to 0.7 µm only, photography includes also the near infrared, spectrometers in the laboratory

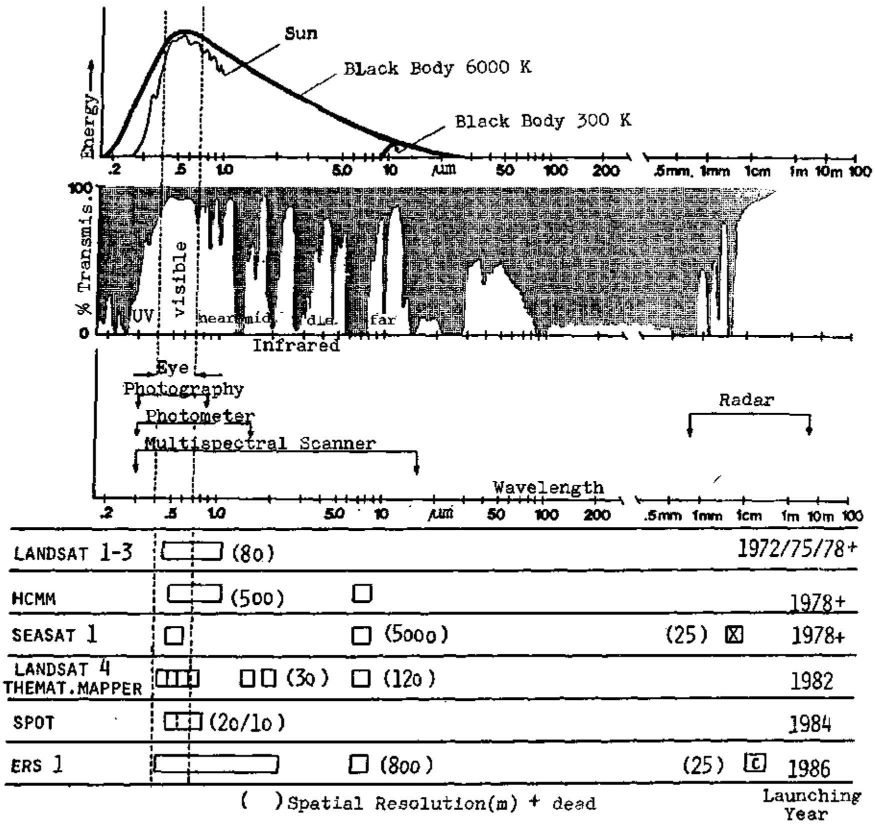


Fig. 1. Spectral ranges of solar and terrestrial emission, transmission and different types of sensors, and their exploitation by different satellites.

usual extend from 0.3 to 1.4 or even to 2.5 and multispectral scanners may be used from aeroplanes or satellites between 0.3 and 15 µm. Such scanners can measure thermal emission from the earth in the range 8 to 12 µm and with shorter waves (visible to mid infrared) the spectral reflectance of soils and plant is measured from direct or indirect solar radiance. In long-wave regions of the spectrum (about 1 mm to 10 m), radar systems are increasingly being used, especially in the tropics and in temperate regions, because they are independent of the weather.

Figure 1 lists satellites used for monitoring earth resources. Two trends are striking:

1. Increasing spectral range and resolution (e.g. Landsat 1-3: 4 bands; Landsat 4: 7 bands), Seasat and the European ERS 1 include radar bands of high spatial resolution.

2. Increasing spatial resolution (Landsat 1-3: 80 m; Landsat 4: 30 m; Spot: 20 or 10 m).
The French Spot satellite can be moved interactively, so that sensing of one scene from different angles will allow detection of height differences up to about 20 m.

Initial processing and image enhancement

Any global (or local) survey system based on quantitative remote sensing must include a variety of modules to process, enhance and classify these data (review by Phillips & Swain, 1978). For monitoring, change-detection modules must be provided too. In Figure 2, a scheme of modules for remote image analysing is presented, Some of their functions have been demonstrated by images in the case studies (Jakob et al., 1982a).

Scan acceptor. Since formats and specifications of data of different scanners and receiving stations vary, a special module to read these different kinds of data has to be implemented.

Geometric corrector. Remote-sensing scenes have to be situated and recorded. The scale and projection must be corrected to a given standard, especially if change detection by overlay of scenes is planned. Sensors, platforms and geometry of the scene may cause distortions, which have to be corrected by internal control data or by external terrain models.

Radiometric corrector. Electronic drifts and differences of detectors in given bands may be adjusted by BIAS and GAIN modules until an optimum of visual interpretability of a scene is given. Noise due to sensor, recording or transmission of data should be removed. Different angles of sun to scene and to sensor cause radiometric distortions, which can partly be standardized by a combined terrain and illumination model (Meier, 1981). A calibration module may help to convert sensor readings to radiation and temperature.

Spectral compositor. The recomposition of the values of spectral bands is often a prerequisite for improved interpretations and classifications of images. The construction of uni- and multidimensional histograms and correlation/covariance matrices are first steps in evaluating relations between spectral bands or object classes of the scene. Eigenvectors and ratios may help to extract effects of special image elements (e.g. 'signatures' of soils, plants or geology). For example, knowledge a priori of spectral signatures of eroded and colluvial soils or of increasing crop covers, as shown in Figure 3 for loamy soils and rape seed in eastern Holstein, can be used to assess erosion effects or soil-plant relationships regionally. Both soils and plants have great influence on the near infrared reflection, but the ratio of short-wave to long-wave reflectance improves discrimination (Jakob et al., 1982).

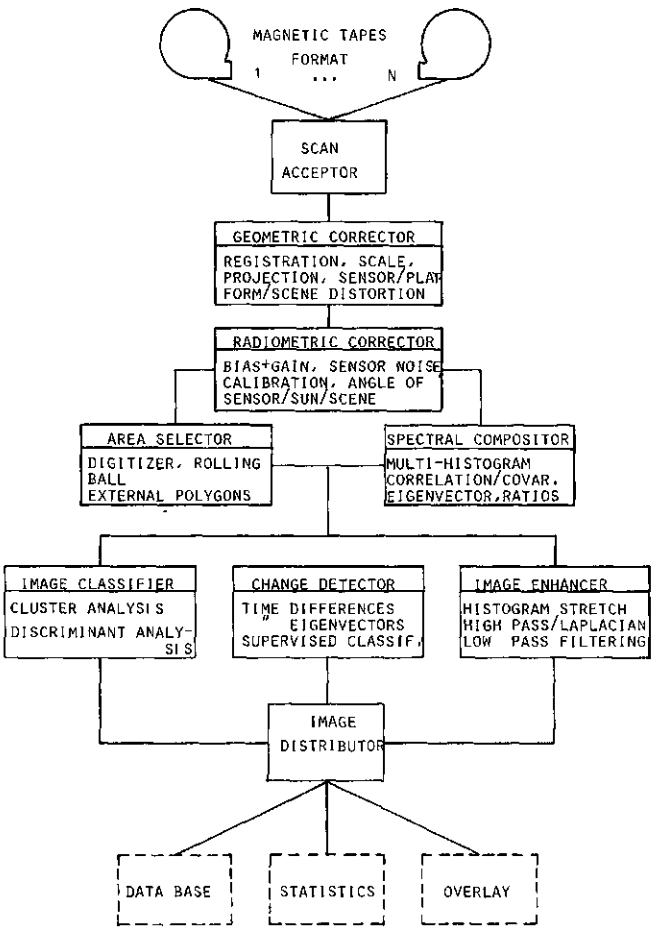


Fig. 2. Scheme of the remote-image analyser module.

Image enhancer. Special enhancement techniques can be used to improve the images. Histogram stretching spreads the original values optimally over the whole range of a spectral band. High-pass and Laplacian filtering techniques sharpen spectral differences at spatial edges of classes (e.g. enhancement of parcelling patterns, Jakob, 1980). Low-pass filtering (running averages) smooth fluctuations in spectral pattern.

Classification and change detection

Image classifier. Spectral classes of images may be classified by a variety of quantitative methods (e.g. Webster, 1977). Many of these routines have been specified for the classification module

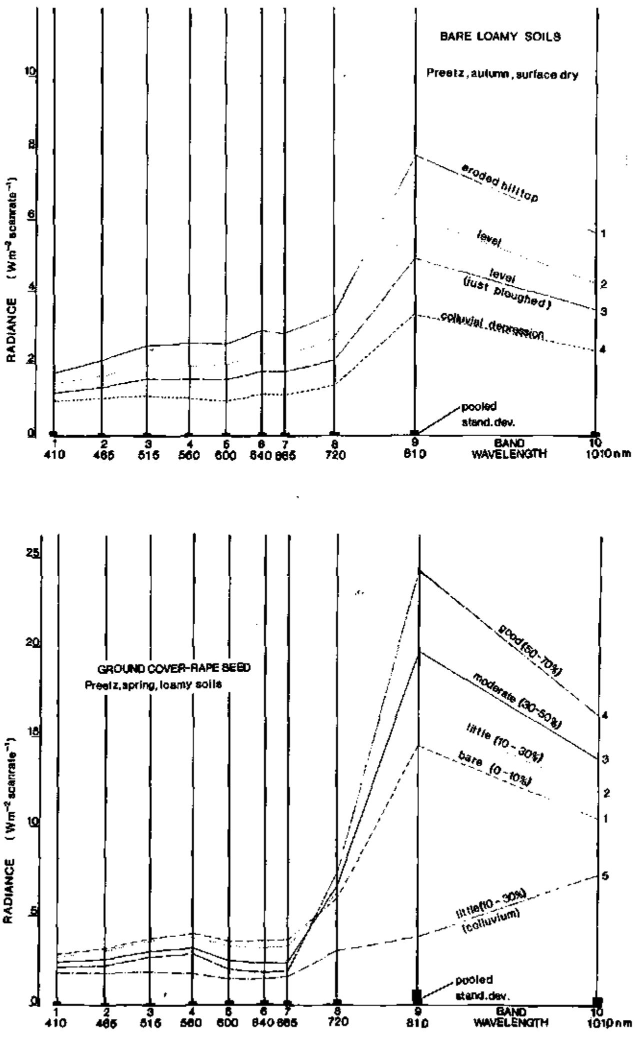


Fig. 3. Spectral signatures of a subsequence of eroded/colluvial soils and rape seed (standardized radiance of 10 bands sensed by an air-borne Bendix-scanner).

of the Global Land/Soil Monitoring System (Bie & Lamp, 1983). They should only be included in the image classifier submodule, if a self-contained remote-image analyser module is available. The automatic (unsupervised) construction of classes by cluster analysis should be used only if classes cannot be determined by the interpreter through limited knowledge about soils and land-uses of the scene. But generally, it is advisable to select and validate training classes by ground-truth surveys and to delineate these classes by the tools mentioned in the area selector module (Fig. 2). Discriminant analysis computes multivariate statistics to assist classification of all other picture elements (pixels).

The correctness of remote-image classifications is closely dependent on the relation of pixels to object sizes, on the complexity of distinctness of object classes and on the spectral resolution of remote sensing. Thus the limited correctness of land-use classifications by former Landsat data, especially in intensive landscapes, should be increased appreciably by the new Landsat or the forthcoming Spot satellite.

Change detector. To improve on single-time analysis, multitemporal classifications can be used. By superimposing two classifications of a Landsat scene from October 1972 and August 1975 Jakob et al. (1982) succeeded in distinguishing different classes of (bare) soils from each other and from residential areas. Also relevant changes in ecological classes (e.g. increase or decrease in pastures, forests or urban areas) were detected. By careful selection and checking of classes and control of natural or man-induced phenology of remote sensing (Jakob & Lamp, 1978) significant differences between seasonal and permanent changes can be observed. Change detection for land and soil monitoring must try to distinguish temporal variations of interest from short-range errors in time and space (Fig. 2).

A special kind of change detection is thermal inertia mapping with far infrared. Emission can be transformed to temperatures with black-body references in the sensor and by physical laws mentioned in the section on Emission, platforms and sensors (though there are problems with background radiation and varying emissivities). Also terrain and illumination effects should be allowed for.

Conclusions

Remote sensing is undoubtedly a major tool for global land and soil monitoring, especially in regions with favourable climatic and natural conditions and with little local information about land and soil. Experiences in intensively cultivated landscapes of north Germany suggest that only advanced methods of multispectral and multitemporal remote sensing can compete at a high-information level. Especially for monitoring land and soil resources based on change detection, geometric and radiometric corrections, and careful image enhancements and supervised classifications must be performed. A prerequisite is the inclusion of point and spatial ground-truth data about land-use and soils by recourse to spatial-resource information systems. The construction and use of such integrated systems should be emphasized in future activities.

References

Bie, S.W. & J. Lamp, 1983. Criteria/soft-hardware for Global Land/ Soil Monitoring System. Version 2, UNEP, Nairobi.

Jakob, J.A., 1980. Landnutzung und Parzellierungsmuster auf Landsat Bildern - eine Hilfe zur Bodenregionalisierung nord-deutscher Landschaften. Internat. Arch. of Photogrammetry Vol. XXIII, Part VII, p. 477-486.

Jakob, J.A. & J. Lamp, 1978. The compilation of agro-phenological crop calendars for remote sensing of cultured landscapes. Proc. Internat. Symp. Observation and Inventory of Earth Resources and Endangered Environment, Freiburg.

Jakob, J.A., W. Kirchhoff & J. Lamp, 1982. Qualitative and quantitative evaluation of airborn-scanner imagery for pedological and agricultural purposes in North Germany. Internat. Geoscience and Remote Sensing Symp. Vol. I, WP 4-4, p. 1-6. München.

Jakob, J.A., J. Lamp & M. Lehner, 1982a. Computer aided soil evaluation methods on Landsat-images in cultured landscapes. Internat. Arch. of ISPRS, Vol. 24-VII/1, p. 397-405. Toulouse.

Meier, R., 1981. Soil moisture detection with microwave- and thermal infrared-sensors in the catchment area Rietholzbach (Switzerland). Remote Sensing Series Vol. 4. Departm. of Geography, Univ. of Zürich.

Phillips, T.L. & P.H. Swain, 1978. Data processing methods and systems. In: P.H. Swain & S.M. Davis (Ed.): Remote Sensing - The quantitative approach. McGraw-Hill, New York.

Webster, R., 1977. Quantitative and numerical methods in soil classification and survey. Clarendon Press, Oxford.

Remotely sensed multiband digital data for soil information systems: problems and prospects in India

L. Venkataratnam, National Remote-Sensing Agency (NRSA), No 4 Sardar Patel Road, Secunderabad-500003, (A.P.) India

Introduction

Remote sensing is the science of collecting and recording of reflected/emitted electromagnetic radiation from the earth's surface and interpreting such data to derive information. For Landsat, reflectance data from four wavelength bands are stored on magnetic tape for ease of handling and classification into meaningful natural resource categories. These digital data on the computer-compatible tape can be converted into analogue or image form for easy visual interpretation. The digital processing of remotely sensed data is a rapid and repeatable process and so very useful in the soil information systems in monitoring the changes in soil surface, crop growth, soil salinity and soil moisture conditions. Since large areas have been affected in India by various degradational processes including salinity and alkalinity, the utility of Landsat multispectral digital data in the soil information systems with reference to soil degradation has been described in this paper, keeping in mind the vast area of the country as well as the barren nature of the salt-affected areas.

Review of the literature

The reflectance or emittance of electromagnetic energy from soil is largely determined by the integrated contribution from each surface grain of the soil and is affected by soil colour, texture, organic matter content, moisture condition, surface roughness, salinity and physical discontinuities occurring on the soil surface. Al Abbas et al. (1972) had shown a relationship between multispectral response and clay content of surface soil samples, indicating the importance of multispectral scanner data and pattern recognition techniques to delineate and map gross textural differences in surface soils. Obakhov & Orlov (1964) and Montgomery (1976) found that all soils they investigated had spectral reflectivity characteristics related to soil colour, while Mathews et al. (1971) indicated the influence of organic matter and iron oxides on the reflectance. Baumgardner et al. (1970) and Kristof & Zachary (1971) used multispectral-scanner (MSS) data to mapping five classes of organic matter content of mineral soils with computerized analytical techniques.

In India, Landsat MSS data have been used to extract information on soils and soil degradation, especially when the soils are devoid of moisture and vegetative cover, with computer-aided multispectral data analysis (Venkataratnam, 1980a,b; 1981). The small-scale colour coded soil maps prepared by using Landsat data are found to be useful in India for regional planning and as a map base for further detailed soil mapping. This is especially true for areas where there is no reliable soil map so far even on a small scale. The experience in India shows that Landsat data can be used to delineate soil units as far as association of soil subgroups, although in some areas even individual soil series can be mapped because of the contrast and abrupt tonal differences in the neighbouring soil units.

Remote sensing of soil degradation

About 175 million hectares of land in India are estimated to be subject to some kind of degradation, leading to reduction in soil productivity. For proper evaluation and planning the reclamative measures, the delineation and mapping of such problem soils is needed, at least on regional scale. Because of the white encrustation of the salts, the saline soils have lighter tone in Landsat imagery compared to neighbouring normal soils. There is an abrupt increase in spectral reflectance of these soils over that of non-saline soils. So one can delineate these soils. Similarly, the eroded soils in a mountainous vegetative region can be easily identified in false-colour imagery, because the eroded areas appear gray and the surrounding vegetation appears red. The areas under 'jhum' or shifting cultivation in mountainous terrain or north-eastern states of India can easily be identified from the Landsat data (NRSA 1977). These 'jhum' areas which are prone to soil erosion appear in lighter tone than neighbouring vegetation-covered terrain, which can be seen in red colour in a false-colour Landsat imagery. It was felt that Landsat multiband digital data would be most suitable for delineating and mapping of these areas under shifting cultivation, especially in the inaccesible areas.

Waterlogged areas can be identified and delineated using the multitemporal Landsat data. In some north Indian plains, the waterlogged areas may have darker tone in a false-colour image, especially when sandy elevated areas are interspersed in these waterlogged areas, which get flooded in some part of the year (NRSA 1979).

Landsat digital data for soil information systems with reference to salinity/alkalinity

It was demonstrated for the first time in India (NRSA 1976) that Landsat MSS digital data would be valuable in giving information on the extent and geographical situation of salt-affected areas. Colour-coded soil maps were prepared later (NRSA 1979) for the entire Haryana State with the Landsat MSS data, marking two classes of soil salinity. It was observed that highly saline soils with electrical conductivity of about 1.2 S/m or more gave higher reflectance than the other soils (Venkataratnam, 1980) because of the white encrustation of soluble salts in all the Landsat wavelength bands. Table 1 shows mean reflectance of different categories

Table 1. Mean reflectance obtained from Bendix M DAS for various categories of soil at different dates.

Category	75-11-27, Band 27:				77-05-02, Band:				81-03-24, Band:			
	4	5	6	7	4	5	6	7	4	5	6	7
1. Highly saline	79.9	115.5	118.8	94.2	123.3	176.8	180.4	134.4	139.2	160.4	167.3	113.6
2. Moderately saline	59.8	80.8	88.5	72.6	102.6	139.2	148.1	115.4	102.0	118.1	127.7	85.8
3. Mobile sandy areas	56.8	89.2	101.2	82.3	100.1	151.1	155.8	117.7	109.6	136.6	143.6	98.7
4. Stabilized sandy areas	47.6	70.0	80.4	68.0	83.1	132.0	139.2	107.2	89.6	110.3	125.2	88.0
5. Vegetation	33.2	37.6	65.1	64.1	69.5	90.5	112.1	91.3	48.5	45.0	112.1	94.8
6. Wet/water-logged areas	43.5	54.2	56.3	44.5	67.6	87.9	106.1	80.7	73.2	80.9	87.7	60.3
7. Water	42.7	45.6	34.3	10.1	56.2	55.5	41.2	16.0	55.7	45.1	30.5	9.8

as obtained from computer-assisted Bendix Multispectral Data Analysis System (M DAS) at different dates of Landsat path. These mean reflectances were obtained after feeding the computer system at least three 'training sets' the coordinates of which were collected from the field observations. The information system to be developed for monitoring salinity in a particular area where extensive reclamation is being undertaken should allow for change in reflectance of the categories in different seasons.

The changes in the spectral values of salt-affected soils can be seen in Figure 1 and 2 for the same area on different dates where the extensive reclamative measures were undertaken to bring this land under cultivation. So very intensive studies have to be undertaken to develop a technique to monitor changes in salinity by a central institution like National Remote-Sensing Agency (NRSA) where databank facilities are available. After collection of the raw data in the form of computer-compatible tapes (CCT) from NRSA Earth Station, NRSA Data Centre can supply the data, either raw or classified, to various user agencies in India in the form of their choice. The remotely sensed data as well as classified outputs could be disseminated easily to various user agencies in India (Fig. 3) when a centralized Soil Data Base could be established. All the classified data available from Landsat can be segmented into 1° standard grid squares compatible with Survey of India topographical maps, where the unit of reference for storing soil and allied data is approximately 1 pixel (o.5 hectare). Thus a unified approach for a total soil information system can be established (Murthy, 1983).

Soil-degradation monitoring system

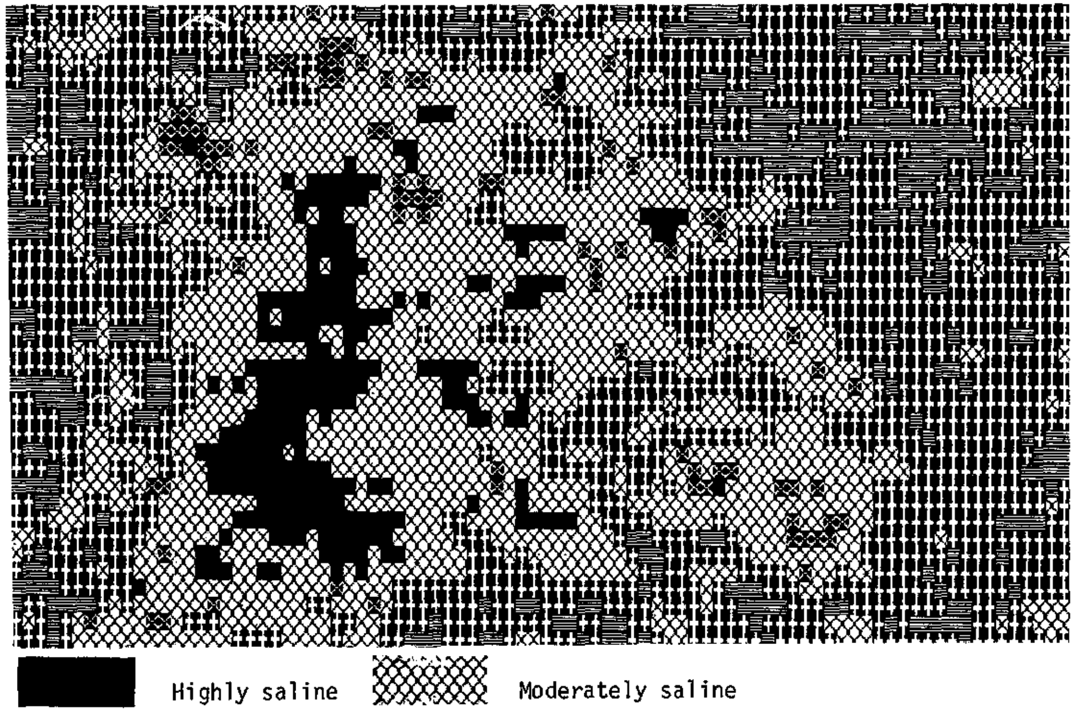
It is absolutely essential to have a soil-degradation monitoring system for a country like India, where millions of hectares of land have been degraded. For such a monitoring system, a soil data bank is needed and the average response values are to be collected together with spectral reflectance curves for each soil unit that can be delineated. The spectral reflectance data should have a wavelength resolution of 5 nm with an accuracy of ± 0.5 % standard error in the spectral reflectance as was also stated by McCloy (1980).

The spectral reflectance of each soil or soil degradation category should be collected for each date of satellite pass in all the seasons for an effective monitoring system, together with intensive ground observations and some soil samples for laboratory analysis over a period. This necessitates the cooperation and coordination of several State and Central agencies with NRSA involving considerable financial obligations. However, once the monitoring system is perfected, season-to-season planning will be easy for both the State and Central agricultural departments.

Highly saline   Moderately saline

Non-soil saline classes are given other symbols

Fig. 1. Categorized soil salinity classes based on Landsat data (Path-Row. No. 158-039, Dt. 2-5-1977).



Non-saline classes are given other symbols

Fig. 2. Categorized soil salinity classes based on Landsat data (Path-Row. No. 158-040, Dt. 24-3-1981).

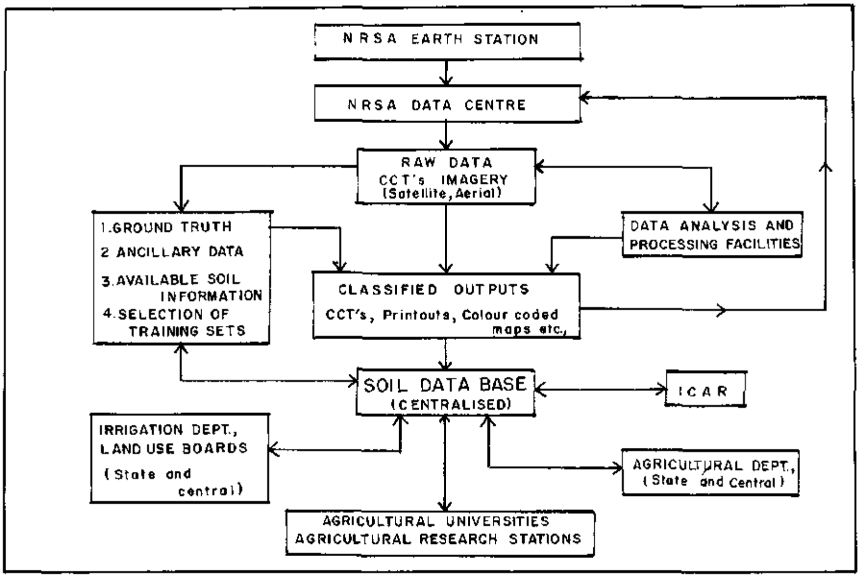


Fig. 3. Suggested soil information system model for India.

Acknowledgment

Thanks are due to Director, NRSA for according permission to present and publish this paper. The technical assistance rendered by various engineers of data analysis facilities is herewith acknowledged.

References

Al abbas, A.H., P.H. Swain & M.F. Baumgardner, 1972. Relating organic matter and clay content to multispectral radiance of soils. Soil Sci. 114(6): 477-485.

Baumgardner, M.F., 1970. Effects of organic matter on multispectral properties of soils. Proc. Indiana Academy of Sci. 79: 413-422.

Kristof, S.J. & A.L. Zachary, 1971. Mapping soil types from multispectral scanner data. Proc. Seventh Internat. Symp. on Remote Sensing of Environment, ERIM, Ann Arbor, USA 3: 2095-2108.

Mathews, H.L., R.L. Cunningham, J.E. Cipra & T.R. West, 1971. Application of multispectral remote sensing to soil survey research in South Eastern Pennsylvania. Soil Sci. Soc. Amer. Proc. 37(1): 88-93.

McCloy, K.R., 1980. Interfacing soil and remotely sensed data types - implications and benefits. Proc. Second Australian Meeting of ISSS working Group on Soil Information Systems, Canberra, Australia, 19-21 Feb. 1980, p. 83-89.

Montgomery, O.L., 1976. An investigation of the relationship between spectral reflectance and the chemical, physical and

genetic characteristics of soil. Ph.D. thesis, Purdue University, Indiana, USA.
Murthy, J.D., 1983. Personal communication. NRSA, 1976. Pilot project for satellite remote sensing survey of parts of Punjab and Haryana. NRSA Abstract Report, Secunderabad (AP) India.
NRSA, 1971. Satellite remote sensing survey of natural resources: Nagaland, NRSA Project Report, Secunderabad (AP) India.
NRSA, 1979. Satellite remote sensing survey of Harayana State. NRSA Project Report, Secunderabad (AP) India.
Obakhov, A.E. & D.S. Orlov, 1964. Spectral reflectivity of the major soil groups and possibility of using diffuse reflections in soil investigation. Soviet Soil Sci. 11: 174-184.
Venkataratnam, L., 1980a. Use of remotely sensed data for soil mapping. Photonirvachak, J. Indian Photo Interpretation and Remote Sensing 8(2): 19-25.
Venkataratnam, L., 1980b. Delineation and mapping of agricultural soil limitations/hazards in arid and semi-arid tropics using Landsat MSS data - An Indian example. Proc. 14th International Symposium on Remote Sensing of Environment, Ann Arbor, Mich., USA, p. 905-914.
Venkataratnam, L., 1981. Delineation and mapping of problem soils with remotely sensed data. Presented in the Interactive Symposium and Workshop on 'Applications of remote sensing for rice production' held at NRSA, Hyderabad (AP) India, 9-13 Sept. 1981.

Towards digital geomorphological and vegetation maps

Wolter Arnberg, Department of Physical Geography, University of Stockholm, Drottninggatan 120, S-10691 Stockholm, Sweden

Abstract

Experiments to mix raster and vector data form part of a project to build up a digital geomorphologic and vegetation database covering 1325 km^2 in the central part of Sweden. The software system is presented and the experience using an Osiris image-scanning machine for cartographic purposes as well as for manual data-gathering is discussed.

Introduction

In the region of the large Lake Siljan, geomorphological and vegetation maps covering 3125 km^2 have been produced by the Department of Physical Geography, University of Stockholm. The project was commissioned by the county administration of Kopparberg. Regular cartographic methods have been applied, but the different vegetation, soil and land forms are kept on separate stripping films in order to simplify scanning of each category.

In an earlier paper (Arnberg, 1981) the design concepts and structure of a geographical software system at the department was outlined. It was emphasized that the data bank could comprise raster as well as vector data.

In this study, the vegetation and geomorphological maps will be entered into the data bank in vector and raster form. This makes it possible to use different resolution of different features of the map. The area-covering symbols will be scanned at low resolution and the roads and boundaries of property units and parcels will be manually traced on a digitizing table at a higher resolution.

This paper will describe a preparatory experiment to transform a vegetation map to digital format. The map sheet is 13 E Vansbro NO according to the topographic map sheet designation.

Scanning of image-separated maps

The category or image-separated sheets are an intermediate step in the map-making process. In brief, the interpreted vegetation boundaries which are drafted on an overlay to the aerial photograph are transferred by a photogrammetric instrument to a plotted map, which is copied to scribing film. The boundary lines are manually scribed and later copied onto stripping films, one film for each vegetation category. In this case 21 films were produced. Different sets of these are then used to produce the printing plates. A problem affecting the scanning process is that the proofreading and correction is made on the printing plates. Another problem may be the minute space in the border lines between different categories. This space causes colour fringes in the printing process and must be masked out by means of a black printed border line.

The stripping films are too big to be scanned by Optronix or Osiris scanners and therefore reduced copies have been used. In this preparatory experiment, the reduction was made with a Hasselblad camera with a 100 Planar lens to 60 mm film of high contrast (Agfa Ortho, developed in Kodak D-19). Normally the stripping films are registered using punched holes. However, to register the scanned images an extra set of register marks have to be introduced. This set of register marks are of a self-adhesive type and stuck on the light-table outside the stripping films. The extra set of register marks are measured on the GTCO digitizing table and the coordinates are expressed in the national grid system ('rikets nät') using the map-frame corner marks, which also are present on the light table. The stripping films are then registered, using corner marks which are visible in the peel-coat, but will not be recorded in the photograph.

The reduced plates are then scanned in an Osiris machine. In this machine, it is possible to access 25 film frames on the film carrier (60 mm film). The linear resolution is 54µm, corresponding to about 0.5 mm in the map sheet. Hence, all areas in the map less than 0.5 mm will be lost, but in the map there are never dots less than 2 mm and seldom linear features less than 1 mm.

The borders between different vegetation or soil types are seldom distinct. Usually there is a gradual transition from one type to another over a distance, often in the order of 10-100 mm. A reasonable resolution of this type of information in a digital image is a picture element (pixel) size corresponding to 25 m x 25 m. This is a fifth of the linear resolution that is available in the printed map. However, roads often need better resolution. The smallest mapped vegetation unit is 1 ha.

To get a properly registered set of digital map layers, the scanning system has to be calibrated. Light fading, specks in the optics, and different sensitivity of the different diodes of the diode array of the Osiris have to be compensated for by scanning without a photographic image in the machine. A set of such 'empty' images are used to calculate the mean intensity of each pixel of an image. These image values are then subtracted from other Osiris-scanned images. The program 'OSIRIS-CALIB' creates the calibration image. The Osiris-scanned images are stored on tape. The tape files are dumped on disk using the program 'OSIRIS-DUMP' which also calibrates the data and reorganizes the image coordinate system to begin in the upper left corner of the image.

The next step is to relate the different images to each other. Then the register marks have to be found in each image layer. These marks - fine crosses - are to be found in almost the same areas of all images. To speed up the search, the program 'MARK-CHASING', has been developed. In this program, an area with a possible mark is searched with a matching algorithm. In this case, the absolute minimum difference between template and image is used. As we also know that we have crosses and that the lines of the crosses are about parallel to columns and lines of the image, the minimum line and column sum is another measure of the position of the register mark in the image. The two algorithms may give different results, and the operator has to judge which coordinates are best. A symbol-coded picture of the register mark is then printed on the line-printer for verification, using the program 'SYMPIC'.

The image line and pixel position is then used to calculate the constants for plane-affine transformation to the national grid system (program AFFIN). The next step is transformation of the images. This is performed by the program 'TRANSFORM-MATRIX' and can be done with 3 methods: nearest neighbour, bilinear interpolation or cubic convolution. In this case, resampling according to nearest neighbour and bilinear interpolation methods have been used. Pixel size is defined as 25 × 25 m on the ground.

Now we have images that can be superimposed to form a map. As the information from 20 of these 21 layers is mutually exclusive, they can be merged into one image layer. This is done with the program 'COND-ASSIGN-PIX'. Even if the photographic film has a definite edge between a black and a transparent area, the slit size of the scanning equipment will cause a gradual transition in the digital image. It is often a delicate problem to define a threshold yielding a binary image in which the areas have neither grown nor shrunk, for instance by light fading or film properties. A threshold half-way between black and white in the middle of the image may not be good in the corners. In each image, a local adaptive thresholding technique is possible, but to merge many grey-tone images, which

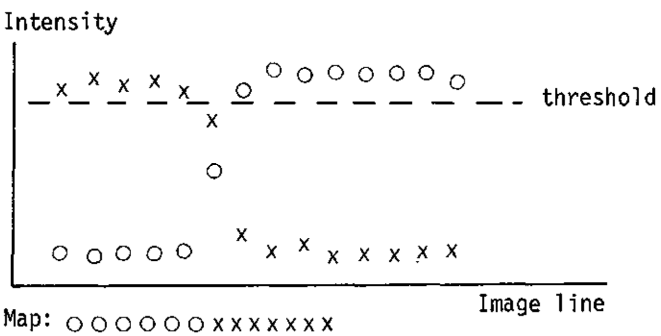


Fig. 1. Principle of the boundary-computing algorithm. Rings and crosses are different map layers.

are inherently binary, into one multicategory grey-level image, the following boundary-computing or adaptive algorithm has been used. Assign all pixels under a user-defined threshold value as candidate for the actual category provided that the actual image value is less than any previously tested. If a pixel is assigned, store the image value in a scratch layer (Fig. 1). Map categories that are not mutually exclusive are stored in separate image layers or added to one layer. The program 'ASSIGN-PIXEL' is used to assign every pixel in the destination image layer according to a user-defined interval in the scanned image. If a vector representation of the thematic map is desired, the multicategory image will be used. This prevents bad definition of the borders and indicates properly which parts or segments of a border line are shared by other categories.

Manual digitizing

Maps are sometimes very complex with different levels of information although 3 symbol types represent three types of information (points, lines and areas). Texts could be connected to these types as well. Also, the symbols present information collectively, since the distribution of symbols over an area gives information about relative position, distribution and structure. In addition, the symbol representing a feature is not absolute, but relative to scale and the purpose of the map. A printed map is always a compromise between available data and an effective communication to the user. A digital map as such does not need to be for both data storage and communication but only for data storage. However, the map-output program should be the communicative part and allow the cartographer to choose, the size of different symbols, their position and generalization. The aim of the digitizing program should be to maintain the resolution of the original data, whether it is a map or an aerial photograph. When digitizing different maps, one problem is the change of a symbol from correct planar extent of a feature to a symbol characterizing one or several occurrences of the feature. This can often be seen in geomorphological maps, where drumlins for example, are mapped if they are large enough, with a symbol representing the true extent of the drumlin. In other areas, a single symbol represents the presence of one or more drumlins.

Digitizing program concept

The coordinate capture is done in a formalized manner in order to structure the data already at input. A map element is the basic map unit and could, for instance, be a border line, coordinates of a symbol or text. Areas or polygons are treated as sets of labelled region boundaries or edge segments, which are latched together at junction points called nodes. Lines are connected in nodes as well. Tracing of lines and boundaries can be made in arbitrary direction and order. After digitizing all border lines are processed to form polygons. This routine also resolves all 'islands' inside a polygon, calculates the proper area and gives the polygon a unique identifier. This identifier is the coordinate of the central point of the polygon (or the nearest point whithin the polygon), and is the key to the user label. Sometimes there are overlapping areas and lines of many types. It is possible to keep 99 separate map 'layers' for different levels of information. However, there are cases when there are transitions between point symbols and mapping of the plan extent of a feature, or we have features that we want to outline but do not want to treat as polygon or line data. Hence, the data type 'outline' is available. This data type could also be used to digitize symbols showing different sizes and orientations.

In an earlier version of the program, the operator had to enter a code connected to the map elements before the coordinate data was captured. In the new program version, this is not needed. This will significantly speed up the digitizing. After any editing, the polygon formation is done. In the older version an automatic check of the name assignments of the polygon segments was done in order to detect missing segments. Now this is done afterwards.

MIDAS program and file structure

The digitizer program is included in a program package called MIDAS - Map Information and Data Analysis System - and handles acquisition, editing, storage and retrieval of georeferenced data. The digitizer program is designed to handle both 2-D and 3-D coordinates from a variety of input devices, for example, digitizers and photogrammetric instruments. It is also possible to use the program in connection with automatic curve-tracing machines. During digitizing, the operator will use a series of programs for different purposes. The first program will be MIDIG - the digitizing program. A preliminary output of the digitized map could then be done with MIPLOT. Erroneous lines or texts are then deleted or changed with the program MIEDIT. When all digitizing has been done, the program MIPOLY is run. This program performs polygon formation and checks thematic codes. After that the map data bank will be accessed through MIMAP, which is a program under development. The MIDAS programs will create and access at least 4 files:

- JOBNM-MEBF:MDAS Map Element Bank File
- JOBNM-EREG:MDAS Map Element Register
- JOBNM-NREG:MDAS Map Node Register
- JOBNM-PRXX:MDAS Map Polygon Register on level XX

JOBNM is the job name, a unique alpha-numeric name for one map sheet. An interface to digital image processing is the program MIPTM, which converts the polygon data into a digital image.

Map output and display

Scanned map data can be displayed in the EBBA system, in a simpler colorgraphic unit called Datacolor, and as hard copies. EBBA is an image memory containing 3 image planes of 8 bits each and 4 binary image (graph) planes. The EBBA system has some image-processing capability of great interest in cartography. Although the size of the image is only 256 × 256 pixels, the principles are well presented. To transfer the scanned and compiled digital map to EBBA, the communication program 'EDUMP' is used. In this preparatory experiment, copies of the 20 stripping films were transferred into 20 discrete gray levels. In the image memory of EBBA, the source data are always filtered through a 'transfer function' before the image is displayed on the monitor. This transfer function is a reference table accessed by the video generator. This means that manipulations of this table will cause contrast and colour changes simultaneously of the whole image. Hence, it is very easy to aggregate classes. Area calculations are performed on the image histogram. Road data can be stored as graph planes. To produce hard copies, this type of reference table must be handled by the output program. The example shown in Figure 2 is produced by the program 'FALSE-COLOR' which is a program to make full false-colour or pseudo-colour coded images with reference tables. The 'FALSE-COLOR' program uses UNIRAS routines. This makes the image processing partly device-independent. The UNIRAS modules may create a raster

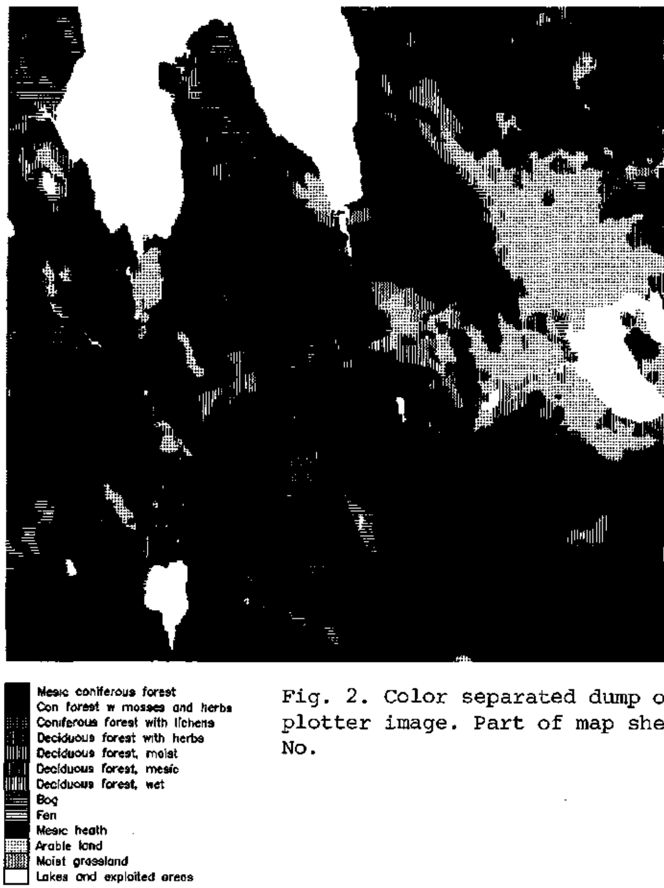


Fig. 2. Color separated dump of an ink jet plotter image. Part of map sheet 13 E Vansbro No.

image on disk, which can be accessed later if data from different sources should be merged. Using different device-dependent 'drivers', the digital image is output. At the moment, drivers for Hertz A4 and Applicon ink-jet plotters, Prism colour impact printer and Trilog 1100 have been implemented. The advantage of using UNIRAS is that the raster (grid) size is defined for a certain output medium, independent of the resolution of the input data. That makes it possible to mix raster and vector data arbitrarily. This is the aim of the program MIMAP. For the present, only the MIPLOT program is running. It produces plots on the Trilog plotter/printer or on the Hewlett Packard 9872 pen plotter.

Results of the scanning experiment

Merging of the digital versions of the scanned photographs showed that some border pixels remained unassigned when using nearest-neighbour resampling method. Presumably this is mainly due to residual errors in the image-transformation process. The nearest-neighbour algorithm is fast but introduces a position error in the resampled image of up to 0.7 pixels. In addition, the residuals of the transformation computations are in the order of 1 pixel. Together, this may cause a slivering effect in the composed image. To avoid this, smoothing of the images can be done. Bilinear interpolation, in which the four pixels surrounding the calculated coordinate are involved in the resampling process, blurs the image slightly and this method, in combination with the boundary-competing algorithm, gives a proper result.

Some data in the vegetation map, for example clear-cut areas in forest, are overlaid with other information. This type of information can be stored in a separate image layer as shown above. However, *having 20* mutually exclusive classes, they will occupy 5 bits of the byte. The three bits left can be used for 3 superimposed layers. To manipulate the reference table more easily, the three least significant bits are used for this purpose. As most images are stored either as binary (bit) or byte images in our system, this will economize the storage.

The scanning of reduced copies of image-separated maps indicates that it is possible to use the Osiris machine. This statement has also some other implications, regardless of digital maps. This method offers an improved and easier way to calculate and display area statistics. At the same time a more age-persistent and less bulky storage medium is used, giving back-up safety. At present, a photoelectrical area-measuring method is used with reduced copies of the stripping films. Scanning the vegetation-map separates and merging gave different results for many categories (Table 1). However, checks of the digital map have not shown any large errors.

The 13 E Vansbro vegetation map sheet will occupy approximate 1 *Mbyte in raster format (1000 x 1000 pixels,* 2 image layers and label). However during processing, each of the layers will occupy half of that much data space. Geometric transformation, using bi-

Table 1. Comparative area estimates of original and scanned digital maps.

	Area (km^2)	
	Original	Digital
Coniferous forest with lichens	25	39
Coniferous forest, mesic to moist	460	390
Coniferous forest with mosses	8	10
Deciduous forest, mesic	7.5	9.6
Deciduous forest with herbs	1	1.3
Thicket and deciduous forest, moist	3	3.9
Deciduous forest, very wet	1,5	1.6
Bog with pines	7.5	9.1
Bog, hummock vegetation, few dwarfed pines	4.5	5.9
Fen, Deciduous or coniferous forest	3	3
Fen, floodplain and similar	2	1.9
Fen with shrubs or lawn-like vegetation	6.5	7.3
Fen, soft carpet-like vegetation	1.5	1.7
Fen, mud bottom vegetation, e.g. flarks	1	2.7
Mesic heath	2.5	2.2
Mesic grassland and arable land	9	12
Moist grassland	< 1	0.8
Lakes	73	116

linear interpolation, will use abaout 1150 CPU sec/layer and merging of the layers to a single image 130 CPU sec/layer on a NORD-100 Ce computer. When the geomorphological map over the same area is scanned, it will be merged with the vegetation map. The volume of data may increase due to data that are not mutually exclusive. It is possible to reduce the volume of data by coding procedures or by raster-to-vector conversion, but so far the amount of data does not seem overwhelming.

Acknowledgment

This study is a part of the project 'Studies of Remote-Sensing Methods' funded by the Swedish Natural Sciences Research Council. Certain aspects of program development and applications have been explored in the projects 'Experiments in Computer-Assisted Remote-Sensing for Forest Survey and Management', 'Remote Sensing Servicing Society' and 'Possibilities of Remote Sensing in Surveys of Natural Resources' funded, respectively, by the Swedish Board for Space Activities, the Swedish Council for Planning and Coordination of Research, and the Swedish Environmental Protection Board. The computer resources have been made available by grants from Knut and Alice Wallenberg's Foundation.

Project manager is Professor Leif Wastenson. Contributions to the MIDAS program package have been made by Mr Göran Alm (MIPTM and parts of MIPOLY and SYMPIC). Mr Hans Hertling is responsible for the implementation of UNIRAS and utilities connected with UNIRAS.

Mr Per Lundblad, Physics IV, Royal Institute of Technology, has assisted in the managing of OSIRIS. Mr Henrik Österlund has made the photographic reductions of the vegetation map and performed the scanning and computer-image processing.

Reference

Arnberg, W., 1981. Integration of map and remote-sensing data - Software framework at the Department of Physical Geography, University of Stockholm. Geografiska Annaler Ser. A 63: 319-324.

List of participants

United Nations

Harvey Croze, United Nations Environment Programme, P.O. Box 30552, Nairobi, Kenya

Bulgaria

Alexander N. Sadovski, N. Poushkarov Institute of Soil Science, 5 Shosse Bankya, BG-1080 Sofia

Denmark

Henrik Breuning Madsen, Arealdatakontoret, Landburgaministeriet, Enghavevej 2, DK-7100 Vejle
Flemming Duus Mathiesen, Arealdatakontoret, Landbrugsministeriet, Enghavevej 2, DK-7100 Vejle
Steen W. Platou, Arealdatakontoret, Landbrugsministeriet, Enghavevej 2, DK-7100 Vejle

Federal Republic of Germany

Wolf Eckelmann, Niedersächsisches Landesamt für Bodenforschung, Postfach 510153, D-3000 Hannover
Jürgen Hatje, Institute of Plant Nutrition and Soil Science, Christian-Albrechts-University, Olshausenstr. 40, D-2300 Kiel
Jürgen Lamp, Institute of Plant Nutrition and Soil Science, Christian-Albrechts-University, Olshausenstr. 40, D-2300, Kiel
Reiner Vinken, Niedersächsisches Landesamt für Bodenforschung, Stillerveg 2, D-3000 Hannover

France

Raymond Van der Driessche, ORSTOM, SSC, 70-74 route d'Aulnay, F-93140 Bondy
Ana Gerschenfeld, Association de Coordination Technique Agricole, 149, Rue de Bercy, F-75595 Paris Cedex 12

India

Lukka Venkataratnam, National Remote Sensing Agency (NRSA), No. 4 Sardar Patel Rd., Secunderabad, 500003, A.P.

Kenya

Victor D'Costa, Soil Science Dept., University of Nairobi, P.O. Box 30197, Nairobi

Mexico

Raul Ponce-Hernandez, p.t. Soil Science Laboratory, Dept. of Agricultural Science, University of Oxford, Parks Rd., Oxford OX1 3PF, England

The Netherlands

Johan Bouma, Netherlands Soil Survey Institute, P.O. Box 98, NL-6700 AB Wageningen

Peter A. Burrough, Dept. of Soil Science and Geology, Agricultural University, P.O. Box 37, NL-6700 AA Wageningen

Ben A. Marsman, Netherlands Soil Survey Institute, P.O. Box 98, NL-6700 AB Wageningen

New Zealand

David Giltrap, Soil Bureau, DSIR, Private Bag, Lower Hutt

Norway

Stein. W. Bie, Norwegian Computing Center, P.O. Box 335 Blindern, Oslo 3

Odd M. Braaten, Jordregisterinstituttet, Postboks 115, N-1430 Ås

Ola Einevoll, Jordregisterinstituttet, Boks 115, N-1430 Ås

Jon Frank, Institutt for jordbunnslaere, Postboks 27, N-1432 Ås-NLH

Arne Grønlund, Jordregisterinstituttet, Boks 115, N-1430 Ås

Jarle Hole, Jordregisterinstituttet, Postboks 115, N-1430 Ås

Torfinn Kummen,Jordregisterinstituttet, Boks 115, N-1430 Ås

Svein Erik Lønnum, SysScan A/S, Boks 131, N-3601 Kongsberg

Eivind Solbakken, Jordregisterinstituttet, Boks 115, N-1430 Ås

Eiliv Steinnes, University of Trondheim, N-7034 Trondheim

Poland

Henryk Kern, Institute of Soil Science and Plant Cultivation, Izabelli 5/8, PL-24100 Puławy

Sweden

Wolter Arnberg, Dept. of Physical Geography, University of Stockholm, Drottninggatan 120, Stockholm

Turkey

Suat Senol, Dept. of Soil Science, Faculty of Agriculture, University of Cukurova, Adana

Union of Soviet Socialist Republics

Yakov Pachepsky, Institute of Soil Science and Photosynthesis, SU-142292, Puschino, Moscow region

United Kingdom

Philip Beckett, Dept. Agric. Science, University of Oxford, Parks Rd., Oxford OX1 3PF, England

David P. Bickmore, David Bickmore Associates, Bletchingdon, Oxford OX5 3DP, England

Donald Mackney, Soil Survey of England and Wales, Rothamsted Experimental Station, Harpenden, Herts. AL5 2JQ, England

Mary Proctor, Soil Survey of England and Wales, Rothamsted Experimental Station, Harpenden, Herts. AL5 2JQ, England

Richard Webster, Soil and Plant Nutrition Department, Rothamsted Experimental Station, Harpenden, Herts. AL5 2JQ, England

United States of America

Gordon L. Decker, Soil Conservation Service, US Dept. of Agriculture, P.O. Box 2890, Washington DC 20013

Venezuela

Simon Arenas, p.t. Soil Science Laboratory, Dept. of Agricultural Science, Parks Rd., Oxford OX1 3PF, England

Juan Comerma, National Center for Agricultural Research, CENIAP, Apdo. 4653, Maracay 2101-A.

Yugoslavia

Suad Alagić, Dept. of Informatics, Faculuty of Electrical Engineering, University of Sarajevo, YU-71000 Sarajevo

Zambia

Cuulu Kalima, Soil Survey Unit, Private Bag 7, Chilanga

Nicholas Mumba, Dept. of Agriculture, Lusaka

Witze Veldkamp, Soil Survey Unit, Private Bag 7, Chilanga